AALBORG UNIVERSITY
DENMARK

STANDARD TITLE PAGE FOR PROJECTS

To be completed by the student(s)

Subjects: ***(tick box)***	Project :	Thesis: X	Written Assignment:

Study programme:	International Marketing	
Semester:	4^{th}	
Exam Title:	Master's Thesis	
Group Number:		
Names + Student Nos of group member(s):	Name(s)	Student Number(s)
	Martin Bobula	20171374
	Jan Petr	20171213
Submission date:	6^{th} of June, 2019	
Project Title /Thesis Title	Designing CRM retention strategies using Deep Learning Customer Churn Prediction model in a subscription-based company	
According to module description, maximum number of characters/words/pages of the paper	150 pages	
Number of characters/words/pages *(Standard page = 2400 characters including Tables and Figures, excluding References, Appendices, Front Page, Table of Contents)*	127 pages	
Supervisor (project/thesis):	Reimer Ivang	

We hereby declare that the work submitted is our own work. We understand that plagiarism is defined as presenting someone else's work as one's own without crediting the original source. We are aware that plagiarism is a serious offense, and that anyone committing it is liable to academic sanctions.

Rules regarding Disciplinary Measures towards Students at Aalborg University:
http://www.plagiarism.aau.dk/Rules+and+Regulations/

Date and signature(s):

____
Signature and date

____
Signature and date

____
Signature and date

____
Signature and date

____
Signature and date

____
Signature and date

____
Signature and date

Designing CRM retention strategies using Deep Learning Customer Churn Prediction model in a subscription-based company

Jan Petr, Martin Bobula

Table of contents

Summary

Creating, maintaining and retaining relationships with a company's customer base is one of the crucial business and marketing tasks in heavily competitive markets such as telecommunications or subscription-based models. The researchers therefore take this thesis as an opportunity to thoroughly explore the topic of Customer Relationship Management, more specifically Customer Churn and Retention Strategies. The topic is explored via systematic literature review from different angles counting business context, degree of CRM individualization, segmentation, selection of key customers, enabling process, employee engagement and performance assessment. Additionally, the contemporary tool of Deep Learning is explored and utilized in order to predict customer churn and yield supplementary marketing insights. As a result, a conceptual framework compositing the existing literature on the topic of CRM Retention strategies is created and later synthesized with the outcome of the Deep Learning Customer Churn Prediction model. This approach allows the researchers to conclude to what extent can the latter enhance the former. The results indicate that the utilization of Deep Learning in predicting customer churn represents an effective way to cope with customer defection. Not only is this approach capable of accurately predicting which customers are going to churn, it can also provide understanding of churn predictors which in combination with churn reasons can yield valuable marketing insights. These results can be supplemented by additional information such as customer lifetime value or proneness to marketing interventions. However, such a model ought not to standalone in the process of the whole CRM Retention strategies development as multitude of other factors such as business context or retention goals come into picture.

Keywords: customer relationship management, retention strategies, customer churn, Deep Learning

Introduction

Deep Learning has become a phenomenon of a great interest in the context of Customer Churn Prediction. This approach to identifying churning customers within a firm's customer base is yielding extremely promising results in the subscription-based business model, more specifically in the telecommunications industry. However, all the important academic papers examining its capabilities are focusing on the topic strictly from the Data Science perspective. As a result, no broader marketing context is supplied to the Deep Learning Customer Churn Prediction models.

The researchers therefore take this thesis as an opportunity to explore the topic of Deep Learning Customer Churn Prediction within the context of CRM retention strategies. The concept of CRM retention strategy is reviewed by the existing literature, and a conceptual framework for the development and implementation of CRM retention strategy is created. Furthermore, the research presents the theoretical background for constructing a Deep Learning Customer Churn model in order to understand how this method can be embraced in subscription-based companies. This information is subsequently used in a case of Taiwanese music streaming company KKBox, where a Deep Learning Customer Churn Prediction model is built, and all the steps of the process are thoroughly described.

The findings from the Deep Learning model show that this approach is an effective tool for predicting customer churn with a high accuracy, however, the complex nature of the technique may negatively affect the interpretability of the model. In the context of CRM retention strategies, the predictions of customer churn on an individual level can be used when deciding which customers should a company target through the retention efforts, though, this decision ought to be based on other factors as well. In addition, the information about predictors of churn yielded by the interpretation of the model can be used as additional insight in understanding reasons behind customer churn. Lastly, the insights about each individual customer contribute to creating a complete picture about each customer which serve as a foundation for individualized CRM retention strategies.

The main contribution of this thesis to the marketing field is rooted in putting the previously well-researched standalone area of Deep Learning Customer Churn Prediction to the context of CRM Retention Strategies. At the same time, novelty approach of creating multi-input Deep Learning architecture is employed where the authors of this thesis combine tabular static data and image recognition representing customer behavior into a single Deep Learning model.

This paper is mostly relevant to marketers and data scientists who can hereby observe the effects of Deep Learning Customer Churn Prediction model in the broader marketing context. Marketers can furthermore utilize the synthesis of the two phenomena from this project for the purpose of building a fully informed CRM retention strategy while employing the most contemporary approach to customer churn prediction. At the same time, subscription-based companies with similar data nature can adopt the customer churn prediction approach researched and empirically examined in this paper in order to increase the predictions' accuracy.

As a result, this paper serves to fill the knowledge gap caused by the fact that the vast majority of academic papers focusing on Customer Churn prediction via Deep Learning focus on the tool autonomously and omit the broader marketing perspective. Furthermore, a novelty approach to building the Deep Learning architecture is presented.

Problem Formulation

Customer Relationship Management

Customer Relationship Management (CRM) is the most profitable and fundamental approach for creating and sustaining relationships with customers (Soltani et al., 2018). Therefore, in the last two decades the area of CRM has become a topic of a thorough marketing research (Foltean et al., 2018). Moreover, companies with more resources for CRM have better financial performance and process capabilities (Keramati et. al, 2008). Furthermore, CRM offers opportunities to use information and data in order to understand customers and later co-create value with these customers (Payne et al., 2005). Payne et al. (2005) describe the concept of CRM as a strategic approach uniting potential of relationship marketing strategies with the help of IT in order to create improved shareholder value.

CRM is also tied with technology, where the utilization of CRM link the technological innovations to their ability to deliver product value to individual customers and to design effective customized communications, as well as to analyze and collect data on customer patterns, interpret customer behavior and develop predictive models (Soltani et al., 2018). Moreover, the participation of Information Technology has a significant and positive influence on the CRM activities (Ko et al., 2008; Soltani et al., 2018). This also means that by investing in the CRM technology, managers are able to improve their CRM capabilities, which subsequently result in an improvement of the business performance (Wang et al., 2012).

The research of CRM can be organized along the customer lifecycle which includes three main stages: customer acquisition, customer development and customer retention strategies (Kamakura et al., 2005). Firstly, customer acquisition extends from the channels the customers use to first access the firm to the promotions that bring them to a firm (Ansari et al., 2004; Kamakura et al., 2005). Secondly, the customer value can can be also increased through appropriate development strategies such as delivering customized products and cross-selling (Ansari et al., 2003; Kamakura et al., 2003; Kamakura et al., 2005). Finally, Kamakura et al. (2005) claims that early detection and prevention of customer churn can also increase the total lifetime of the customer base, if the subsequent efforts are focused on the retention of key customers.

Customer churn as one of the marketing's focal points

Consequently, the collective customer base sum of Customer Lifetime Values equals Customer Equity and creates an approximation for the company's value. It is therefore of utmost importance to not only acquire new customers and cross-sell the current ones, but also to devote resources to retain the customers who already form the firm's current base (Gupta et al., 2006). Studying customer churn has therefore become one of the marketing's hot topics. Sure enough, various markets across the world are becoming increasingly more saturated, with more and more customers swapping their registered services between competing companies. Companies have thus realized that they need to focus their marketing efforts on optimizing customer retention next to focusing on customer acquisition (Hadden et al., 2007).

Dawkins et al. (1990) in their Harvard Business Review article claimed that an increase of 5% in customer retention results in the improvement in firm's profitability of up to 85% (Dawkins et al., 1990). Moreover, a report from Lindgreen et al. (2000) claims that acquiring a new customer could be up to 10 times more expensive than retaining a current one. Many other researchers (Grönroos, 1991; Coviello et al., 2002; Buttle, 2004) confirm that aim to improve customer retention has been one of the prominent priorities for the companies which focus on building relationships with their customers. Furthermore, the authors (Dawkins et al., 1990; Reichheld, 1996; Buttle, 2004) agree on the possible economic advantages brought by focusing on averting customer churn (Ang et al., 2006). On the other hand, although the report by Dawkins et al. (1990) report has become heavily influential (given the number of academic papers referencing to it), Sharp (2010) argues against the case in his empirical-evidence-based book How Brands Grow. Firstly, he notes that the case was not built upon any empirical data but a simple thought experiment. Secondly, the authors completely left out the cost linked to increasing customer retention expecting that averting customer churn comes at zero cost. Thirdly, the marketing scholar builds on years of extensive empirical research across many different industries and types of companies and says that customer defection is an inevitable marketing force that befalls all brands. The companies should therefore rather focus on customer acquisition at the expense of customer retention (Sharp, 2010).

Traditionally, three ways of focusing on retaining customers were introduced by Rosenberg et al. (1984): "customer portfolio analysis, customer-retaining marketing mix and reorganization for customer retention" (Rosenberg et al., 1984). Customer portfolio analysis consists of customer's history of purchases either for each product or across the whole company. Companies then may split their customer base into first-time purchases, repeat purchases and churning customers and devote particular marketing efforts toward each segment. Consequently, company should measure customer satisfaction via surveys (Rosenberg et al., 1984). The second approach, customer-retaining marketing mix, suggests dual separate marketing strategies - partly for customer acquisition, partly for customer retention. This strategy takes into account that customers now expect more from the company than the initial offer that engaged them in the first place. Therefore, Rosenberg et al. (1984) suggest "product extras, reinforcing promotions, sales force connections, specialized distribution, and post-purchase communication," to be the key parts of the retention strategy (Rosenberg et al., 1984). Lastly, as the customer churn is often understood as an inevitable force, and is therefore fairly overlooked, the third retention approach suggests a proper measurement and attention to customer churn rates. This strategy may involve an executive position devoted solely to managing customer retention (Rosenberg et al., 1984).

It is clear that the historical approach to customer retention involved customer surveys, aggregated metrics and statistics regarding the entirety of a company's customer base. On the other hand, with the increased possibilities in fetching data, and its consequent storage and processing, contemporary approach to customer retention differs heavily. The information about each customer's demographics, purchase trends or service usage allows companies utilizing

Artificial Intelligence to assess the likelihood of churning on the level of an individual customer, together with probable churn drivers (Rosenberg et al., 1984; Zhu et al., 2017).

Artificial Intelligence in the world of Marketing

In the recent years, the area of Artificial Intelligence, including Machine Learning, Deep Learning and Neural Networks, has made a tremendous progress which opened new opportunities for the academic research, as well as the application in many fields, including business activities and firm development (Li et al., 2018).

Marketing intelligence, which emphasizes the marketing related aspects of business intelligence, has traditionally relied on market surveys to understand consumer behavior and to improve product design (Fan et al., 2015). Nowadays, in marketing intelligence, data (big data, by extension) relevant to a company's markets is collected and then processed into valuable insight which is used as an additional information in the processes of decision making (Hedin et al., 2014, Fan et al., 2015). The term 'Big data' describes large volumes of complex, high velocity, and variable data that require advanced technologies and techniques to enable the capture, storage, distribution, management, and analysis of the information (TechAmerica Foundation's Federal Big Data Commission, 2012).

The different methods that can be applied to discover marketing intelligence are based on the characteristics of the available data. When the developed model is based on a single data source, the outcome may provide only limited insights, and therefore potentially lead to biased decision making. However, by integrating heterogeneous information from different sources, the outcomes provide a wider view of the area which generates more accurate marketing intelligence. Since the integration of big data from multiple sources is not a trivial task, new applications and methods for effective big data management in the context of marketing intelligence is being explored (Fan et al., 2015).

Over the years, researchers find more and more methods and applications of big data analytics in the marketing field. Amongst the areas that made the most noticeable progress in the field of marketing are: recommendations, segmentation and customer churn prediction (Fan et al., 2015; Gordini et al. 2016). Recommendations are the main domain in the e-commerce context and streaming services where based on previous choices or customer's attributes, each customer receives different offers as what to buy next or what to watch or listen to next (Dias et al., 2008; Fan et al., 2015; Fernández-García et al., 2019). In the process of segmentation, the big data analytics enables profiling each individual customer such that the most suitable products can be marketed to the most appropriate individual instead at the right time, instead of identifying and segmenting groups of similar customers (Abbasoğlu et al., 2013; Fan et al., 2015). Lastly, slicing a company's customer base as narrowly as to the level of an individual, and consequently predicting future churning customers is an area where the marketing-related research within Artificial Intelligence has registered significant advancement (De Caigny et al., 2018).

Machine Learning and Customer churn

Thanks to the nature of the data available, both academic and scientific research with regard to customer retention has progressed the furthest in one business area in particular: subscription-based type of service (more specifically in the telecommunications industry). That is, generally because companies in this industry have abundant information of day-by-day customer activity such as local/international call records, short messages, voice mail, demographics, financial details, and other usage behavior of the customers. This has created an opportunity for Machine Learning to develop predictive modeling techniques to spot trends and patterns in behavior and predict Customer Churn (Amin et al. 2019; Zhu et al., 2017).

Various Machine Learning techniques have been previously used for Customer Churn prediction (e.g. Support Vector Machine, Decision trees and Logistic Regression) (Amin et al., 2017). One of the most advanced ways of predicting Customer Churn is by using the technique of Deep Neural networks. The research focusing on this topic has showed that the Deep Neural Networks perform with high precision when it comes to predicting customer churn. In fact, one of the first researches was undertaken by Castanedo et al. (2014), who have developed Deep Neural Network architecture in order to predict customer churn in a mobile telecommunication network.

The research showed that the architecture outperformed the previously used approach of predicting customer churn by over 4% in the accuracy of the prediction (Castanedo et al., 2014).

One of the more recent uses of Deep Learning was done by Microsoft which has developed a Customer Churn prediction model using Deep Learning for their cloud computing service Azure. The data scientists at Microsoft have developed a model which employs both static information about customers and daily logs regarding the usage of the subscription-based service. Especially the latter part of the input, behavioral data, represents a massive leap forward for predictive models. Not only is the company able to predict which customer is likely to churn, the model is capable of estimating the churn drivers as well. This way, the model sets the company up for tailor-made marketing interventions toward the churning user (Zhu et al., 2017).

Nevertheless, even though the application of Deep Neural Networks for customer churn prediction has been around only for the last few years, the research of the application of Deep Neural Networks is not as explored as other methods predicting customer churn (e.g. Support Vector Machine and Decision trees).

Deriving marketing insights based on Customer Churn Prediction

Customer churn has been tackled from two different angles in previous research. On the one hand, researchers focus on improving customer churn prediction models in which more complex models are being developed and proposed in order to boost the predictive performance (Verbeke et al., 2012). On the other hand, researchers want to understand what drives customer churn and defined important reasons behind why customers churn such as customer satisfaction (Gustafsson et al. 2005).

That being said, the ability to generate valuable marketing insight from the customer churn prediction and to leverage this information into the marketing efforts of a company is not a trivial task. This ability effectively depends on the maturity level of the technology and business teams, capabilities they develop, as well as the strategies they adopt (Arora et al., 2015). However, even though the concept of customer churn prediction via Machine Learning is a trending topic amongst companies and researchers, most of the research has been undertaken for the intention to find the most effective and accurate way of predicting customer churn. Yet, the understanding of how to obtain valuable information and use the prediction for creating marketing strategies and actions is arguably just as important.

One of few studies addressing the managerial and marketing implications of customer churn prediction is a research conducted by by Gordini et al. (2016), who developed a churn prediction model tailored for B2B e-commerce industry by testing the forecasting capability of a new model. The authors of this study suggest several main managerial implications and areas which have to be addressed such as the consequences of identifying a churner and non-churner on the company' retention strategies or identifying the riskiest customer segments in terms of churn and focusing the efforts on these customers to potentially save money (Gordini et al., 2016). The authors of this study suggest that marketing managers can decide to develop tailor-made marketing programs to incentive customers to remain with the firm, where these programs will presumably reduce the likelihood of churn. The Marketing programs can consist of three main marketing strategies: 1) subscription management strategy, 2) long-lasting management strategy, and 3) complaints management strategy (Gordini et al., 2016).

Problem Statement

The phenomenon of big data has found its way into the business and marketing where companies try to use it for better understanding of marketing issues and competitive advantage. The customer churn prediction is one of the most researched and trending areas of using big data analytics, where due to the advancements of research and usability of Deep Learning, the Deep Neural Networks represent highly effective way of predicting customer churn. In order to predict customer churn by using big data analytics, certain structure and nature of the data is required. Therefore, the application and research of these methods have been mainly the focus of the firms in telecommunication industry. The business model of these companies is a form of subscription-based business, where also the additional application of customer churn prediction can be found.

The research of customer churn prediction by using Deep Learning has mostly focused on developing the models with the highest accuracy of prediction. However, there is a missing research regarding the implementation of models predicting customer churn into the customer retention strategies.

Due to the aforementioned benefits of both customer retention and customer churn prediction as well as the seeming gaps in the knowledge, the authors will devote this master thesis to examining the topic of Retention strategies as a part of Customer Relationship Management with the enhancement of Deep Learning Customer Churn Prediction model. In order to answer the mentioned knowledge gap, the objective of the research is to examine the following topic:

"Designing CRM retention strategies using Deep Learning Customer Churn Prediction model in a subscription-based company"

Research questions

1. How to understand the relation between customer churn prediction and CRM retention strategy?
2. What are the most prominent churn drivers and which customers is it beneficial to intervene?
3. How can subscription-based companies utilize Deep Learning in customer churn prediction?
4. Can Deep Learning Customer Churn Prediction enhance traditional methods of CRM retention strategy?

Philosophy of Science

The chapter of *Philosophy of Science* describes the research design of the thesis. In order to conduct a research, the philosophical standpoints of the researchers ought to be explained, together with the overall approach to the research. Based on these assumptions, the choices for methodology and techniques are illustrated. Furthermore, the approach to the process of literature review is thoroughly explained to show how the knowledge of the researched problem was inquired.

Based on the different approaches to constructing the methodology of research, Kuada (2012) presents 4 levels, which cover all the steps required to describe the research design of a study. To keep a consistent structure throughout the whole chapter of *Philosophy of Science* the following 4 levels will be discussed:

1) Philosophical and Theoretical Level (Ontology),
2) Epistemological Level (Epistemology),
3) Methodological Approach (Methodology),
4) Methods and Techniques (Kuada, 2012).

Before describing each of the 4 levels, the paradigm of the research is presented. The concept of paradigm is a cluster of understandings of what kind of phenomenon is being studied, what sorts of questions are useful to ask about the phenomenon, how the researchers should undertake their research and how the results should be interpreted (Kuhn, 1970; Bryman, 1988). The paradigm serves as a summary of researchers' assumptions regarding to ontology, epistemology, views on human nature and methodology (Kuada, 2012).

Paradigm

The classification system by Burrell and Morgan (1979) is one of the most influential in the organizational studies and business research (Kuada, 2012; Bryman et al, 2011). It has an importance in understanding the ontological and epistemological foundations of the business research (Bryman et al., 2011) and identifying four different paradigmatic positions: the functionalist paradigm, the interpretive paradigm, the radical humanist paradigm and the radical structuralist paradigm (Burrell et al., 1979).

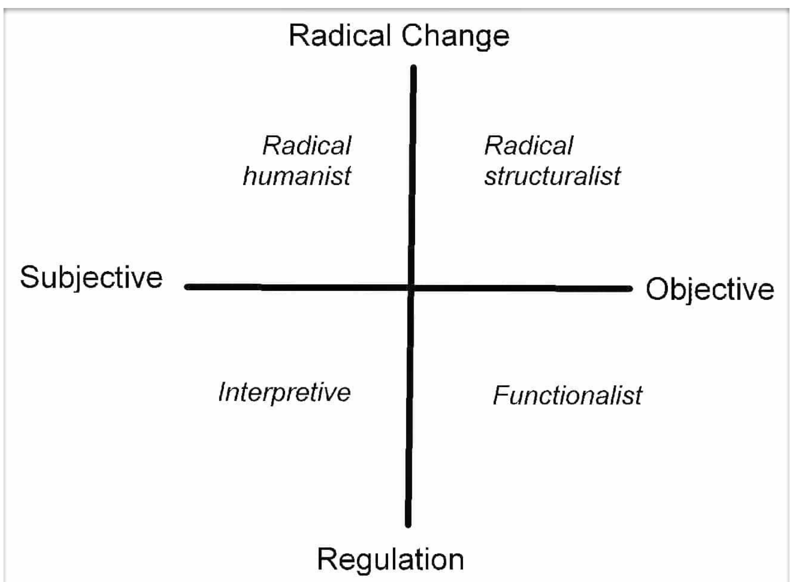


Figure 1: *Burrell and Morgan's Four Paradigms Model for the analysis of Social Theory, Adopted from Burrell and Morgan, 1979.*

The sociology of radical change (the upper half in Figure 1.) is concerned with structural conflict and its purpose is to make judgements about the state of organizations in the business research (Bryman et al., 2011; Burrell et al., 1979). On the other hand, the sociology of regulation (the lower half in Figure 1.) is concerned with status quo and social order and describes what happens in organizations in the area of business research (Bryman et al., 2011; Burrell et al., 1979). Furthermore, the concern of sociology of regulation is on the individual level while the sociology of the radical change is on the society level. Therefore, the research aligns with the paradigm of social regulation. Moreover, the objective of this study does not attempt to make any judgements of what is happening in the organizations, rather just suggests improvements (Bryman et al., 2011; Burrell et al., 1979).

Between the two paradigms of regulation, the research inclines to the functionalist paradigm. This paradigm is described as a combination of objectivity and order and it is a dominant paradigm in the studies of organizations, being based on a problem-solving orientation (Burrell et al., 1979; Bryman et al., 2011; Kuada, 2012). The social issues can be viewed as value free and objective and the researchers can distance themselves from the research. The functionalist paradigm also aligns with the ontological and epistemological assumptions of the researchers and the approach they take when conducting the research.

The chosen paradigm sets up scope to the standpoints in the following steps of the philosophy of science. However, the paradigm is also based on the ontological and epistemological foundations, therefore, these steps are closely connected and have an influence on each other.

Ontology

The purpose of Ontology is to describe the nature of what there is to know about the world (Snape et al., 2003) or what the researcher seeks to know (Kuada, 2012). The key ontological question researchers face is whether the reality exists independently of human interpretations and there is one common reality or reality does not exist independently of human interpretations and there are multiple context-specific realities (Snape et al., 2003; Burrell et al., 1979). The perception of researchers' reality serves as foundation for what is considered as "truth" and how the knowledge about this "truth" should be acquired (Kuada, 2012).

In the ontological decision, researchers assume the pragmatist approach, meaning the objectives of the investigation and the nature of research determine what view of reality is chosen (Kuada, 2012). In the problem formulation the researchers illustrate the objectives of the proposed study based on the current knowledge of the researched field. The aim of this study is to present solution to designing CRM retention strategies by using the technology of Deep Learning. In line with this formulation of the problem, the researchers adopt the objectivist view of reality, which implies that social phenomena and their meanings confront the researchers as external facts that are independent of social subjects and that are beyond their reach or influence (Bryman et al., 2011). This choice is in line with the nature of research conducted before, where the studies of designing customer retention strategy are primarily researched from the objectivist perspective. Although the individual strategies of companies differ in order to fit their business environment, the steps and decisions to design a successful retention strategy are presented objectively and can be generalized. Furthermore, the objectivist view fits to the other element of the research - the Deep Learning area. The key aspect of Deep Learning is based on calculations and layers which are not designed by human engineers, instead they are learned from data by using a general-purpose learning procedure (LeCun, 2015). Thus, the area shall be viewed as objective, existing independently of human interpretations.

However, choosing objectivist view entails aspects that might limit certain scopes of the project, particularly in the research of designing retention strategies. Due to the complexity of the concept of designing strategies, it is nearly impossible to take all factors influencing the strategy development into consideration. Furthermore, many studies of the topic are conducted by qualitative research creating the issue for researchers to generalize their findings across other organizations.

Epistemology

Epistemology is closely tied to ontology and describes the nature of knowledge and the ways of knowing (Kuada, 2012; Snape et al., 2003). Snape at al. (2003) identify three main issues regarding epistemology based on debate in social research: (1) what is the relationship between the researcher and the researched; (2) what is conceived as a truth; and (3) what are the ways in which knowledge is acquired (Snape et al., 2003).

Similarly to ontology, the researchers adopt the objectivist approach to the epistemology, also known as the positivist approach. This standpoint pursuits to conduct the research as objective as possible which is in line with the aim of the research described in ontology. This approach suggests that researchers can be "value free" and independent from the subject that is being researched (Snape et al., 2003). Moreover, based on the previous research of the studied area, researchers believe there is a match between observations of the world and independent reality, which aligns with the positivist approach as well (Snape et al., 2003). The nature of knowledge in the area of Deep Learning can be generally considered as objective since the nature of the data which is collected by algorithms recording online behavior of the customers goes outside of researchers' reach. Furthermore, the Deep Learning model predicting customer churn which processes the collected data makes mathematical calculations which are not influenced by subjective choices of the researchers. Even though creating of the Deep Learning architecture is based on the data available and capabilities of the researchers, the impact of these interactions influences only the accuracy of the model and not the structure of the outcome of the model. In regard to acquiring the knowledge in this project, the researchers choose the approach of induction where the findings of the Deep Learning model serve as the evidence for generating conclusion.

The choice of positivist approach also raises the question whether researchers can be completely objective since the conducted research will always be influenced by assumptions and biases of the researchers. Moreover, the same issue can be drawn with the observations that being studied, which are always being interpreted by researchers to a certain degree. Since the researchers are aware of their biases, the methodology and methods for the research are chosen to help withdraw and minimize these biases.

Methodology

Methodology describes the overall approach to the research and the reasons for choosing specific methods in the process of a research (Kuada, 2012). The choice of methodological approach is connected to and based on the ontological and epistemological assumptions of the researchers

(Kuada, 2012). The strategy of the research can be generally split between quantitative and qualitative research. The choice of objectivist ontology and positivist epistemology is usually connected the quantitative research which uses deduction as principal orientation to the role of theory in the relation to the research (Bryman et al., 2011). However, Silverman (1985) argues that certain quantification of findings from a qualitative research can help to uncover the generality of the studied problem (Bryman et al., 2011).

When choosing the research design, the researchers need to take into consideration how the knowledge can be acquired, as well as the limitations and feasibility of the research. In order to acquire the knowledge and understanding of Deep Learning models predicting customer churn, the researchers need to thoroughly understand and examine this technique. Therefore, the authors choose case study as a research design for the thesis. This choice allows the researchers for a thorough examination of the technique and capabilities and feasibility of the research. Even though qualitative approach in the form of a case study is not standardly used in the objectivist research, in certain situations this approach is viable. Lee et al. (2007) argue that due to the capacity of case studies to thoroughly analyze the dynamics in single setting, the potential for better understanding of organizational phenomena can be provided, in comparison to statistical analysis (Lee et al., 2007).

Yin (2003) defines three uses for case studies: (1) exploratory, (2) descriptive and (3) explanatory. However, Lee et al. (2007) points out that majority of these uses are best understood as poor relations to positivistic, quantitative research. The exploratory case studies tend to be conducted as preliminary research in advance of wide-scale surveys to map out the themes for the subsequent research. The descriptive case studies are often used to expand on trends and themes already discovered by survey research. Lastly, the explanatory case seeks to derive a detailed understanding of a particular phenomenon where the case is not seen as additional to more quantitative methods (Lee et al., 2007). It is the explanatory case that is adopted by the researchers in order achieve the previously mentioned aims of the project, and it is in line with the positivistic ontological and epistemological choice. Furthermore, in line with the fundamental premise of explanatory case study, the researchers are adding to the existing literature by conducting research on their own.

Naturally, the choice of using cases study brings drawbacks and challenges to the research as well. Especially in the approach this research has adapted, the generalizability of findings across subscription-based companies is a major challenge. However, Yin (1994) claims that the crucial question is not whether or not the findings can be generalized to a wider universe, but how well the researcher generates theory out of the findings.

The following sections describe how the findings of a DL CCP model can be utilized in the process of designing CRM retention and how can a subscription-based company adopt this approach. Since Deep Learning is a technique which has found its application only in the recent years, the already existing research of the topic is limited as well. Therefore, based on the latest know-how of the field, the research focuses on the process of developing a Deep Learning model predicting customer churn to understand what are the benefits such an approach can yield. Due to the complicated nature of Deep Learning processes, this approach enables researchers to recognize the reliability of this technique and the advantages and disadvantages that come with it. Subsequently, these findings are used to find out what are the implications in the context of designing CRM retention strategies. Although, the research consists of only a single case, the case serves as example to understand the capabilities of the Deep Learning model predicting customer churn in subscription-based company. It is noteworthy, that the architecture of the model is built and limited by the researchers' knowledge and data available. However, by studying how such model can be developed, the researchers describe which steps are conducted in order to develop a the specific model used in the research. Once the functional Deep Learning Customer Churn Prediction model is developed, the researchers gather all the information the model can provide. These findings are are synthesized with the already existing theories of designing CRM retention strategy which are acquired by literature review. The influence of the Deep Learning model on respective steps of CRM retention strategy is discussed and presented, which leads to reaching the objective of the research.

The findings of the research can be accomplished by a single case as long as the researchers are aware of the limitations of the case and how to generalize this case in a broader perspective. The necessary knowledge for this awareness is acquired in the literature review of researched area.

Methods and Techniques

The level of methods and techniques illustrates the data collection methods and techniques used in the research. The methods and techniques are chosen on the foundations of the previous 3 levels: ontology, epistemology and methodological approach (Kuada, 2012).

As described in the step of methodological approach, in order to reach the research objective, it is imperative to develop a Deep Learning model predicting a customer churn, to find out what insights it can yield. By conducting this step, it is possible to show how can be the technique of Deep Learning used in the process of designing CRM retention strategies. This means the data in the research do not present the answers needed for conclusions but instead serve as a tool enabling creation of the Deep Learning model and consequent exploration of its capabilities to deliver marketing insights. In order to assemble the data needed for such a research, the data is not collected, but instead it is directly fetched from Kaggle. Kaggle is an online community of data scientists, with accessible data sets for purposes of data science and Machine Learning (Usmani, 2017). The data used in this project consist of the demographic data, transaction data and behavioral data of customers of a subscription-based company music streaming company KKBox (see the case description in the *Analysis* chapter). The data are collected by the company by quantitative nonparticipant observation, where the data is registered by the system monitoring online behavior and transactions of the customers. On top of that, the researchers have at their disposal customers' demographics information. In order to develop a functional Deep Learning Customer Churn Prediction model, several steps and techniques have to be implemented.

The entire following process of constructing a Deep Learning customer churn model was done in the programming language Python, which is one of the most used and capable programming languages for the data science purposes. The first step is the data exploration where the data is examined in order to find the problematic characteristics inside the datasets. These characteristics can negatively influence the process of fetching the data into the later created Deep Learning model. Keeping the findings of data exploration in mind, the data is later preprocessed which includes imputing missing values, handling categorical variables, feature scaling, padding and feature engineering. The final two steps of data preprocessing are handling class imbalance and splitting into training and testing dataset. The former increases the training possibilities of the model while the latter is crucial as Deep Learning and Machine Learning in general work on the basis of finding and learning traits from previous information. All these steps are prerequisite to enable the researchers build the Deep Learning architecture.

Once the data is shaped into the structure which can be parsed into the Deep Learning Customer Churn Prediction model, the architecture of the model is constructed. On top of the model predicting customer churn, researchers use LIME library within the Python language for interpretation of the model. Due to the struggles with the model described in the *Analysis* chapter, the researchers were in touch with the author of the model, Marco Tulio Ribeiro from the University of Washington. The purpose of the interpretation package is to show based on which variables the model classified customer as 'churner' or 'non-churner'. The interpretation in this context represents a series of features and their respective weights contributing to each individual classification. Therefore, the interpretation in this sense does not contradict the positivistic standpoint of the research.

Lastly, the outcomes of the model are incorporated into the process of designing CRM retention strategies. The knowledge of how to designing CRM retention strategies is acquired by reviewing the already existing literature. Further, the research attempts to complete the aspects needed for developing customer retention strategies by using Deep Learning prediction model, in order to showcase the process in real-market scenario. However, due to the limitations of the available data, the research is able to cover only some of these steps and only to a certain degree.

Literature Review Process

The role of literature review is to gather the knowledge and information about what is already known in the researched area; how was the previous research approached; what are the inconsistencies in the literature and what are the strengths and weaknesses in the theories (Kuada, 2012; Bryman et al., 2011). Due to the chosen research design of the thesis, literature review has an important role in order to synthesize the contemporary phenomena of Deep Learning algorithms with the traditional marketing concepts of developing customer retention strategies. Therefore, the literature aims to conceptualize the traditional marketing concepts of designing CRM retention strategies that have been research for the past decades, but also set up foundations for constructing a Deep Learning customer churn model.

To match the ontological and epistemological choices, the authors choose the approach of systematic literature review. The research involving systematic literature review is argued to be a foundation for strongly evidence-based approaches (Tranfield et al. 2003; Bryman et al., 2011) and is considered to be useful in the areas of decision making involving conflicting opinions of how things are done (Bryman et al., 2011). Systematic reviews provide the most important practical implications, aim to minimise the bias of researchers and are characterized as objective, systematic, replicable and transparent (Siddaway, 2014). In order to answer the research questions and to undertake the research with the positivistic approach of the researchers, a thorough and objective literature review is required and therefore the systematic literature review is the most fitting. However, it is noteworthy that the systematic approach cannot be completely conducted in the review of the Deep Learning area. Due to the fact that this field has emerged only recently, majority of the research is conducted in a case-specific setting. Furthermore, most of the research is scattered around different data science specialized forums and communities, therefore searching in established databases is not possible. Nonetheless, the conducted systematic review followed five stages presented by Siddaway (2014) who describes detailed guide of how to conduct a systematic review. The respective stages of the systematic review are described in the following section, while the funneling process of selecting studies for the theoretical background of the research, is displayed in the Table 1 and Table 2 at the end of this chapter.

First stage of systematic literature review is scoping, which consists of formulating the research questions and clarifying if a systematic review has already been done in the researched area (Siddaway, 2014). The formulation of research questions is described in the chapter of *Problem Formulation* where the initial scoping of researched area was realized. The formulated research questions are:

1. How to understand the relation between customer churn prediction and CRM retention strategy?
2. What are the most prominent churn drivers and which customers is it beneficial to intervene?
3. How can subscription-based companies utilize Deep Learning in customer churn prediction?
4. Can Deep Learning Customer Churn Prediction enhance traditional methods of CRM retention strategy?

Once the research questions are formulated, researchers need to look whether a thorough systematic literature review of the researched area has already been done, which can help them to effectively gain the relevant knowledge (Siddaway, 2014). In regard to the literature review of designing customer retention strategies, selection of key customers to target for retention and drivers of customer churn, the researchers used the article '*In pursuit of enhanced customer retention management'* by Ascarza et al. (2017). On top of systematic research, the researchers have conducted, Ascarza et al. (2017) provided researchers with a review, key issues and different approaches in the customer retention management, which serve as additional information about the already existing research.

Second stage in conducting of systematic literature review is planning. At first researchers need to decide on the search terms which are based on the above-mentioned research questions (Siddaway, 2014). The search terms of the literature review part focusing on the *designing customer retention strategies, selection of key customers to target for retention* and *drivers of customer churn* was targeted on the following terms: 'CRM retention', 'customer retention strategy', 'retention strategy subscription', 'retention targeting', 'retention targeting subscription', 'retention customer selection', 'churn drivers' and 'churn reasons'. For the part of literature review focused on the development and capabilities of the Deep Learning model were: 'Deep Learning prediction, 'Deep Learning customer churn analysis', 'Deep Learning interpretability' and 'Deep Learning customer churn subscription'. These keywords are selected since they focus on the essential parts of the research questions, where they enable researchers to find the most relevant research. Furthermore, synonyms and combinations of the right terms help to cover more of the relevant studies. The initial search of the selected keywords led to 4,189 studies found, where the distribution of these studies across different databases is presented in Table 1.

Afterwards, Siddaway (2014) instructs to choose the inclusion and exclusion criteria for the review with the justification of these chosen criteria (Siddaway, 2014). Due to the extensive nature of the concept of strategy, only the studies directly referring to the designing retention strategies were included. In regard to scoping the research only to the subscription-based companies, the majority of studies presenting frameworks to developing CRM retention strategies present them in a generalizable way, where if the framework is followed, the strategy is designed to fit the context of the company. Therefore, studies presenting such frameworks, regardless of the industry, do not have to be excluded. However, the literature referring to selection of customers to target for retention is including only studies which take into account the capability of predicting customer churn as it is the fundamental contribution of the Deep Learning method. Since the customer churn prediction is mainly a focus of a telecommunication industry, all the included studies are based in the subscription-based setting. The literature of drivers of customer churn includes only studies which directly present reasons behind customer churn and studies referring to implications of these reasons. Among the studies are also included studies from non-subscription-based context, however these studies aimed to find the general reasons of customer churn, regardless of the industry or business models and therefore can be used to gain additional information about drivers of churn. Once these criteria were implemented on the initial search of the selected keywords, the number of studies were minimized to 278 studies (shown in Table 2), which was a feasible size for conducting the next steps of systematic review.

As the last part of planning, the Siddaway (2014) suggests that the review should be documented in record keeping system which helps to document the findings from the researched literature as well as to describe the decision making of the researchers regarding specific papers. The researchers keep records of the literature review in an excel sheet consisting of name of the research, authors, year of the research, methods used, issues, key findings and philosophical standpoints for each research which was reviewed. This helps the researchers to organize the reviewed studies.

Third stage of the guide presented by Siddaway (2014) is the identification which starts by selecting at least two relevant electronic databases (Siddaway, 2014). To find the relevant literature for the research, the researchers chose the following databases: AAU Primo, Google Scholar, JSTOR, ScienceDirect and SpringerLink. Once the searching process in the chosen databases is performed, an additional search is required in order to not miss any relevant work (Siddaway, 2014). The additional search was based on the references from the studied articles, where the relevant references were examined and used in the literature review if they were relevant to the research. This was a crucial step in obtaining additional relevant literature, where due to the needed inclusion criteria, several relevant studies might have been missed.

Screening and eligibility are the last two steps of the systematic literature review and have purpose of searching through large number of found researches and studying eligible articles to extract relevant information for the research (Siddaway, 2014). By organizing the different studies in an excel sheet, the authors effectively screened through key findings and abstracts of found studies to select the relevant researchers for the researched area. After the relevant researches were studied, and the relevant information was added to the literature review. Once again, the significant references found in these studies were examined and added to the literature review if they contributed to this research. The process of filtering studies by the steps of screening and studying relevant references is shown in Table 2.

Once the systematic literature review was conducted, the synthesis and conceptualization of the presented literature is completed with a conceptual framework as an outcome. This approach of conceptualization is chosen to visualize the relationships between the findings from literature as well as to describe the process of designing CRM retention strategies. The framework prioritizes findings which are supported by number of studies and findings which come from the most reliable and objective and sources. Furthermore, the framework is constructed as a process of designing and implementing CRM retention strategy, since it covers all the aspects influencing CRM retention strategy. In the *Analysis* chapter this form of framework enables to show the reach of Deep Learning customer churn model into different steps of designing CRM retention strategy.

The following Table 1. illustrates the number of studies found across the chosen databases by the initial search of selected keywords. Table 2. shows the filtering process of the initial search through the above-mentioned steps of: inclusion and exclusion criteria, screening and relevant references. After implementing the inclusion and exclusion criteria, several studies had repeated across the different databases, therefore only combined number of unique studies is presented in the Table 2.

Initial search of the selected keywords across the databases	Number of results
AAU Primo	1,621
Google Scholar	1,958
JSTOR	108
ScienceDirect	269
SpringerLink	233
All databases combined	4,189

Table 1: Database search result of selected keywords

Filtering method for selecting relevant studies	Number of studies after using a filtering step
Using inclusion and exclusion criteria on the initial search of selected keywords	278
Screening through the selected studies and selecting the studies that are relevant	46
Selecting relevant papers by screening through the key references of the selected studies	+ 11
The final number of studies used in the theoretical background of the thesis	57

Table 2: Filtering process of the database search result

As the Table 2 shows, the final number of studies in the *Theoretical background* of the thesis is 57, where the overview of the selected studies can be found in the *Appendix* of the thesis.

Quality of the research

Since the aim of a research is to provide trustworthy findings and information to the studied area, it is necessary that the research is designed with certain degree of quality. The quality of the research is commonly evaluated by the criteria of reliability and validity.

Reliability

The reliability is concerned with the question whether the procedures of the study are repeatable by others, and if followed, the same results would be reached (Yin, 2003). In regard to the data used in the research, the data is acquired from Kaggle, which is considered as trustworthy source of reliable data. The methods used to preprocess the data and develop the Deep Learning model predicting customer churn are repeatable if the process in the study is followed. This is due to the reason that these processes are objective in the sense that the data is processed by mathematical calculations which always result in the same outcome. The preprocessing of data follow general standardized steps, as well as constructing the Deep Learning model architecture is based on the foundations of reliable and successful Deep Learning models.

In the part of the thesis studying customer retention strategy, the key findings are found in the literature. The systematic approach to the literature review provides high reliability of this process, where the comprehensiveness of the review is thoroughly described in the prior chapter *Literature Review Process*. The developed theoretical conceptual framework is based on the foundations of already existing framework, where adjustments are made for clearer visualisation of certain steps that are influenced by the Deep Learning model. Furthermore, this framework includes only steps backed up by reliable literature and the different adaptations of the CRM retention strategy framework are addressed.

Validity

The concern of validity is the integrity of the conclusions that are generated from the research, and the concept of validity consists of several layers (Bryman et al., 2011). The layer of construct validity establishes the correct measures for the scope of the thesis that is being studied (Yin, 2003). In order to examine how can the Deep Learning Customer Churn Prediction enhance the process of CRM retention strategies, the research select the approach of constructing the Deep Learning model in the research. This method helps to understand and showcase what is the process and outcome of this technique. Furthermore, this method thoroughly describes the steps that have to be undertaken, as well as the options and obstacles in the process of construction. The results of the research represent the outcome of what such a model can provide.

The internal validity concerns the question of whether a conclusion that incorporates a causal relationship between two or more variables is correct (Bryman et al., 2011). In regard to the concept of designing CRM retention strategies, the systematic literature review helps the authors to understand the different approaches and perspectives to the studied area. By understanding these different approaches, the researchers are able to implement all the relevant variables and relationships to the self-developed framework of designing CRM retention strategies. Furthermore, the limitations of the scope of the research always highlight if there may be any causal relationships which are not being addressed.

On the other hand, the external validity regards the issue of whether the results of a study can be generalized beyond the context of the specific research (Bryman et al., 2011). Although, the choice of case study as a research design weakens the generalizability of the research, the conclusions can be used in different cases and scenarios. As the research provides explanatory analysis to the technique of Deep Learning Customer Churn Prediction, the limitations of this case are addressed. This allows the researchers to generate outcomes which describe the Deep Learning technique objectively and not only as a prediction technique for a single case.

Theoretical background

The purpose of theoretical background is to acquire relevant knowledge for this research. The selected areas that are reviewed are based on the research questions of the thesis, where the following literature review consists of two parts: (1) customer retention strategy and (2) Deep Learning customer churn model. The literature review of customer retention is described by chapters: *Designing customer retention strategies*, *Selection of key customers to target for retention* and *Drivers of customer churn*. The review of these chapters outlines relevant studies and approaches to the researched area where the findings of these studies are conceptualized at the end. The theoretical background of Deep Learning customer churn model provides an introduction to Deep Learning, lists the most frequent Neural Networks used in customer churn and describes in detail two cases which later serve as foundation for the analysis. Furthermore, the chapter gives an insight into the interpretability of Deep Learning networks and outlines two important data preprocessing prerequisites for training such a model - handling imbalanced class distribution and splitting the dataset into training and testing subsets.

Designing customer retention strategies

The following literature review describes different studies addressing and contributing to the development of customer retention strategy. With the introduction of customer retention management, number of studies investigated factors which need to be taken into consideration when developing of CRM strategies. On top of that, certain studies proposed frameworks for the development process of such strategies. Apart from developing CRM retention strategies of subscription-based businesses, the review presents relevant researches of the topic in different industries and business models.

One of the first papers discussing key issues of customer retention and its strategic aspects is the article ‘Customer retention: a potentially potent marketing management strategy’ written by Ahmad et al. (2001). Ahmad et al. (2001) present different potential customer retention strategies for different situations based on previous theories. Further strategies are proposed based on suggestions of experts, as well as observed strategies in companies. As a result of different approaches to customer retention, Ahmad et al. (2001) highlight that the main factor of developing retention strategies is to make them appropriate to their business context.

The paper by Weinstein (2002) examines how companies should develop customer retention initiatives and focus to maximize their long-term customer value and describes how segmentation can help in developing retention strategy. Specifically, Weinstein (2002) suggests that by classifying customers based on their usage variety and frequency, companies are able to create profitable and effective strategies for their customers. Based on the case of plastics manufacturer, Weinstein (2002) distinguishes between (1) heavy users, which require the most attention and individualized marketing plans should be targeting these users; (2) medium users, where strategies for growing this base must be developed; and (3) light users, which should be served in cost-efficient programmes.

There are several studies researching the implementation of CRM in companies. Based on their results, researchers of these studies suggest factors and steps which are beneficial for companies to implement in the process of strategy development. One of the studies conducted in order to research CRM implementation was conducted by Yim et al. (2005). In this study researchers identified activities for effective CRM implementation and later on investigated their effect on customer satisfaction, customer retention and sales growth. The results of collected survey data of 1,223 service firms showed that (1) customer satisfaction; (2) organizing around CRM; and (3) managing knowledge; positively affect customer retention (Yim et al., 2005). Furthermore, this research showed that focusing on key customers also positively affects customer retention, although the results show that the effect exists only through customer satisfaction (Yim et al., 2005).

The failing implementation of CRM by companies was also the motivation for the research conducted by Ryals (2005). The main findings presented in this study demonstrate that the implementation of CRM activities delivers greater profits. Using calculations of the lifetime value of customers in two longitudinal case studies, the research finds that customer management strategies change as more is discovered about the value of the customer (Ryals, 2005). These changes lead to better firm performance. The contribution of this study shows that implementation of CRM is beneficial for companies and a relatively straightforward analysis of the value of the customer can make a real difference (Ryals, 2005). Furthermore, the results of the research propose that the important issue is not customer retention, but profitable customer retention where managers should prioritize retention of larger customers (Ryals, 2005). The larger customers create higher total value which is closely linked to the lifetime revenues of customers (Ryals, 2005). Lastly, the results of the two studies show that the higher total value of larger customers is achieved despite the higher costs of these customers (Ryals, 2005).

Strategies needed for companies to improve the customer retention are presented by Zineldin (2006) who researched the relationship between CRM, customer loyalty and quality. This research is based on the previous studies describing customer satisfaction and loyalty as key elements of profitability, where the more satisfied customer is, the more loyal the customer is which conclusively makes the relationship with customers stronger (Zineldin, 2006). By designing research model to measure satisfaction and loyalty, Zineldin (2006) claims that the retention strategy composed of quality and tactical programs must be designed and afterwards implemented. The presented strategies which can improve customer retention are: (1) measuring customer retention rates over time; (2) analyzing the reasons why customers are leaving the company; (3) focusing on the most profitable customers; and (4) focusing attention on employees to ensure that the offered product or service meets the requirements of the targeted customers (Zineldin, 2006).

Kumar et al. (2005) studied the use of customer-level marketing strategy and its impact on the performance of companies. Due to the demand of customers for personalization and customization of products and services, the strategies of firms have to adapt to the current business environment (Kumar et al., 2005). In this study the researchers develop a framework evolving firm strategy based on customer-level approach. This framework (Figure 2.) attempts to create an interface between the marketing and finance disciplines by connecting the value of each customer, determined by evaluating the lifetime value of the customer to the firm, with the performance of the firm, using seven customer-level marketing tactics as differentiating factors (Kumar et al., 2005).

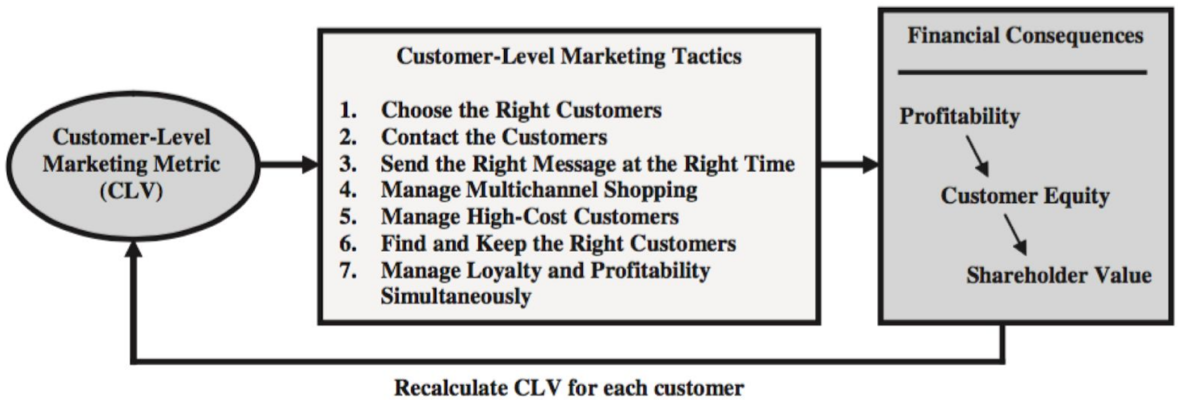


Figure 2: Evolving Firm Strategy, Kumar et al., 2005.

The researchers back up this strategy by its successful implementation in several B2C and B2B companies. However, in order for a company to implement a customer-level strategy, the data collection needs to have 4 characteristics: (1) the data have to be at the customer-level; (2) the data have to contain thorough information of transaction; (3) the data have to stretch across sufficient time period; and lastly (4) the data need to contain information describing when and what kind of marketing intervention previously happened (Kumar et al., 2005).

One of the most extensive researches on the topic of CRM has been done by Payne and Frow. Over the years, their research focused on different perspectives of CRM strategies. In their early work Payne et al. (1999) present a framework for segmented service strategy consisting of 4 steps which is based on a survey of service providers. The steps consist of: (1) defining the market structure; (2) segmenting the customer base and determining the segment value; (3) identifying segments' service needs; and (4) implementation of segmented service strategy (Payne et al., 1999). The researchers highlight that although companies understand the importance of customer retention, only very few measure the economic value of their retention strategies (Payne et al., 1999). Therefore, following the framework enables companies to select appropriate budgets to various customer segments based on the projected lifetime profitability (Payne et al., 1999).

In their further research Payne et al. (2005) create a conceptual framework for CRM strategy which is based on previous literature of process selection criteria for business and marketing processes, and workshop of experienced CRM practitioners who had reviewed these criteria. As a result, the framework consists of 5 processes: (1) the strategy development process; (2) the value creation process; (3) the multichannel integration process; (4) the information management process; and (5) the performance assessment process (Payne et al., 2005). The interaction between the different processes works in both directions with the feedback loops, due to the iterative nature of the development of CRM strategy and flow of information (Payne et al., 2005). In regard to the step of strategy development, the organization is required to focus on both, the business strategy and the customer strategy (Payne et al., 2005). The business strategy must be considered in order to articulate the vision of the company, as well as to review the competitive and industry environment (Payne et al., 2005). The customer strategy is typically designed in marketing departments and consists of identifying the appropriate segmentation of the customers which leads to deciding whether macro, micro or one-to-one segmentation is appropriate (Payne et al., 2005).

This framework was reviewed and adjusted by their research in the following year, where they identify four critical implementation components of a successful CRM programme and examine them in the context of five key cross-functional CRM processes (Payne et al., 2006). The initial model, and the development of further versions of it were refined by interactions with experienced CRM executives (Payne et al., 2006).

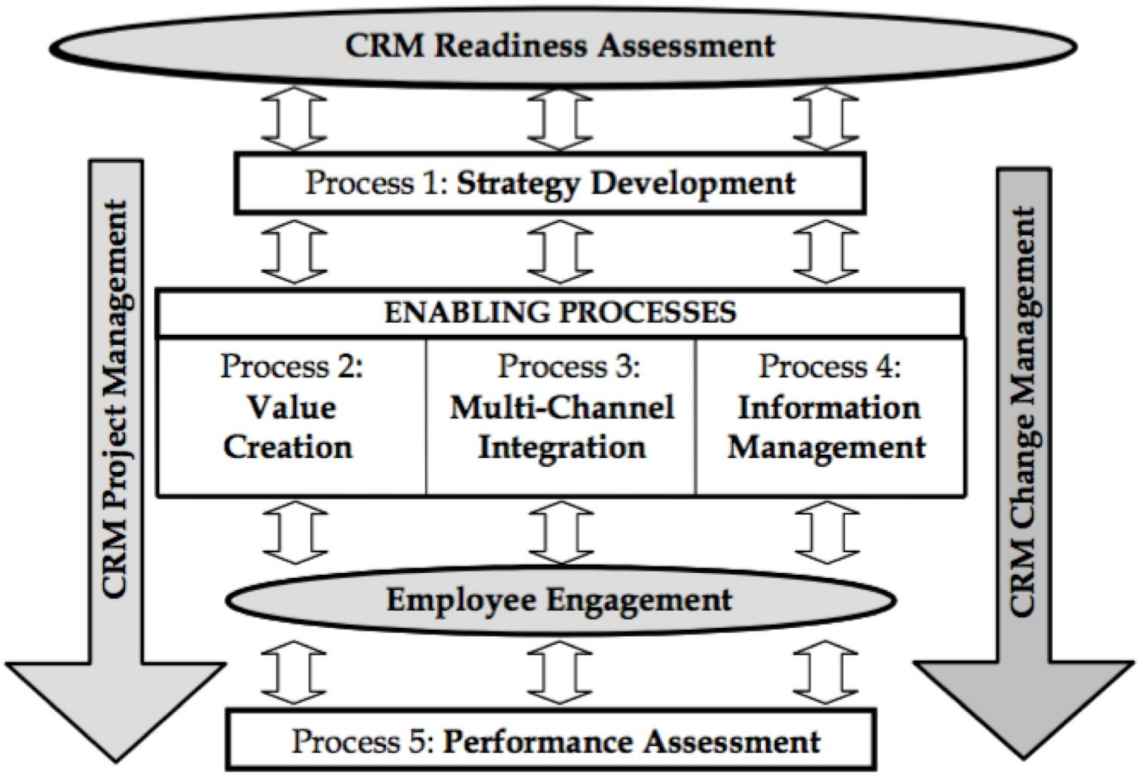


Figure 3: CRM Strategy and Implementation Model, Payne et al., 2006.

In the part of strategy development, on the top of the suggestions from their previous research, they highlight the priority of business strategy and customer strategy alignment, especially when they are being developed in different parts of business (Payne et al., 2006).

The strategic context of CRM is later addressed by Frow et al. (2009) in their work 'Customer Relationship Management: A Strategic Perspective'. Due to the high prevalence of CRM failure in the literature, this study presents range of important strategic issues that are to be considered by companies (Frow et al., 2009). The presented issues (Figure 4.) are based on the interviews with experienced executives working in different sectors and areas of CRM (Frow et al., 2009).

Customer segments: Who are the existing and potential customers? Which forms of segmentation are most appropriate, rather than easiest to undertake? What are the major segments? What are the opportunities for micro-segmentation, one-to-one marketing and mass customization?

Customer relationships: What kinds of relationship does the company have or want to have with customers? How retainable are the customers? How do we 'remember' customers? Is customer communication fed back into the business so it can relate to customers on a one-to-one basis?

Product/service involvement and complexity of customer purchasing behavior: Who constitutes the customer decision-making unit? How are products/services purchased? How important are they to customers?

Company's profile: Where does the company fit within the industry structure? What is their strategic intent? What are the organization's resources and competences?

Stage of industry evolution: What are the current state and likely future changes in industry structure?

Competitors: What is the nature of competitors? How do they compete? How will new competitors evolve in the future? Are there new entrants on the horizon that are not hindered by the same legacy architecture? Are there new strategic alliances that may disrupt the market?

Channels of distribution: What is the current and future role of different distribution channels? What are different opportunities that exist for disintermediation or reintermediation? What opportunities exist for new forms of electronic distribution and delivery?

Information technology platform: What is the appropriate information technology platform and software to serve present and future customer and corporate needs?

Figure 4: CRM Strategic Issues to Consider, Frow et al., 2009.

In the same study Frow et al. (2009) present CRM Strategy Matrix (Figure 5.), which illustrates how can organizations develop CRM strategies that are appropriate to their industry context, degree of competitive intensity and stage of CRM sophistication. Based on how complete the customer information is and the degree of possible customer relationship individualization, the matrix shows which type of customer relationship is appropriate (Frow et al., 2009).

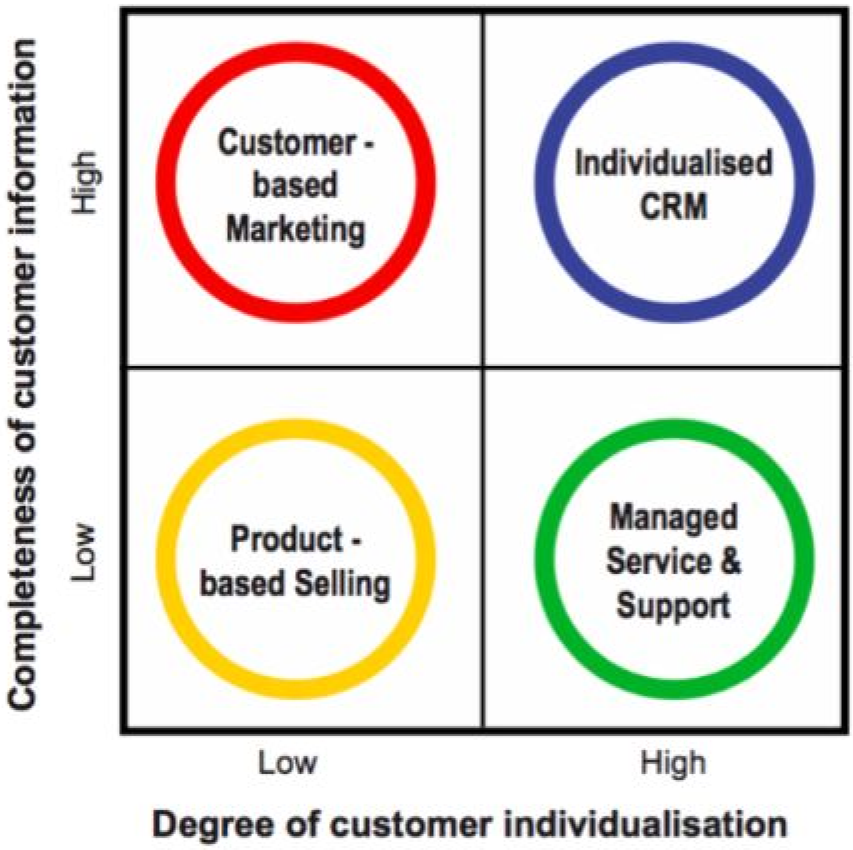


Figure 5: CRM Strategy Matrix, Frow et al., 2009.

However, Frow et al. (2009) highlight several issues which need to be considered when choosing the appropriate type of customer relationship. Even though many firms may consider shifting towards customer relationship with higher customer individualization by using more complete information, firstly they should take into account the trade-off of cost and benefits of developing these individual relationships (Frow et al., 2009). Furthermore, the information on the individual level of a customer requires to cover all aspects of customer behavior, as well as the process of acquiring and maintaining this accurate and complete data is expensive and time consuming (Frow et al., 2009). Finally, this approach should be chosen only when customers have sufficient profit potential and the customers want to be engaged in an individualized relationship with an organization (Frow et al., 2009).

Designing retention strategies in subscription-based environment

As the presented literature shows, the different frameworks to designing customer retention strategies are presented in generalizable form. Therefore, the decisions in respective steps of strategy development have to be appropriately designed to fit with the context and objectives of the company. In regard to the research of customer retention in subscription-based companies, the majority of the research takes place in the telecommunication industry. This is due to the fact that the ability to retain customers in the highly competitive and increasingly saturated telecommunications market enables companies to make considerable profit (Jeng et al., 2012).

Gerpott et al. (2001) researched customer retention, loyalty and satisfaction in the German telecommunication market. The research in this study was conducted via telephone surveys of 684 customers of network operators in the country (Gerpott et al., 2001). The findings of the research suggest that when a telecommunication operator wants to develop a retention strategy, they should not rely only on the general indicators of customer satisfaction and customer loyalty when they try to determine the threat of churning of customers (Gerpott et al., 2001). However, the mobile network operators should rather aim to improve their measurements of perceptions the customers have about the core services they offer. There are two key elements describing the degree of customer retention, which are also tools for improving the relationship between customer and provider: (1) the customer's perception of the prices being fair; and (2) the perception of customer of the functional benefit of the offered services (Gerpott et al., 2001).

Xevelonakis (2005) in his article presents a framework for developing effective customer strategies based on customer behavior, customer risk and customer profitability. Similarly to other studies, the motivation for this research was the incapability of telecommunications companies to successfully benefit from their CRM efforts. Xevelonakis (2005) claims that in order to develop an effective customer retention strategy, companies have to: (1) understand the customer behavior; (2) define the probability of churn; and (3) identify customer profitability. Based on these steps, companies are able to build customer portfolio and develop proper strategy. These steps are further used in the framework of developing effective customer strategies consisting of overall seven steps:

1) Analyze internal and external data,
2) Build appropriate customer clusters,
3) Identify customer profitability,
4) Identify the propensity to churn,
5) Create a customer portfolio,
6) Apply customer profitability and propensity to churn to the identified clusters to design loyalty campaigns,
7) Design and carry out profitable campaigns (Xevelonakis, 2005).

Selection of key customers to target for retention

As mentioned in the previous chapter, one of the crucial parts of customer churn strategy is identifying the customers which should be targeted with the retention efforts. Based on the several researches showing the profitability of customer retention, firms started to shift their efforts into identifying which customers are at the highest risk of churning and embarked on targeting their efforts towards them (Blattberg et al., 2008; Ascarza, 2017). However, the customers with the highest probability of churning do not have to be the ones which are most beneficial for the company to focus. The following literature review shows different views on the targeting of customers for customer retention purposes.

The research by Neslin et al. (2006) analyses different customer churn prediction models. Neslin et al. (2006) created a framework for profitability of a single churn management campaign based on which they estimate the most successful methods and what are the most important factors when creating a customer churn predictive model. However, this study is based on the assumptions that the best way to manage customer churn to predict the customers' probability to churn and afterwards target the ones whose probability of churning is the highest. This way a firm is able to focus its efforts on those customers who are truly at the risk to churn, and potentially save money by not targeting the customers who are not at this risk (Neslin et al., 2006).

Blattberg et al. (2008) built on the work of Neslin et al. (2006) and created profitability framework (Figure 6.) for proactive targeted churn management. Blattberg et al. (2008) developed this framework on the similar arguments as Neslin et al. (2006), where he claims that only the 10% of customers in the segment of high churn are the “real” churners.

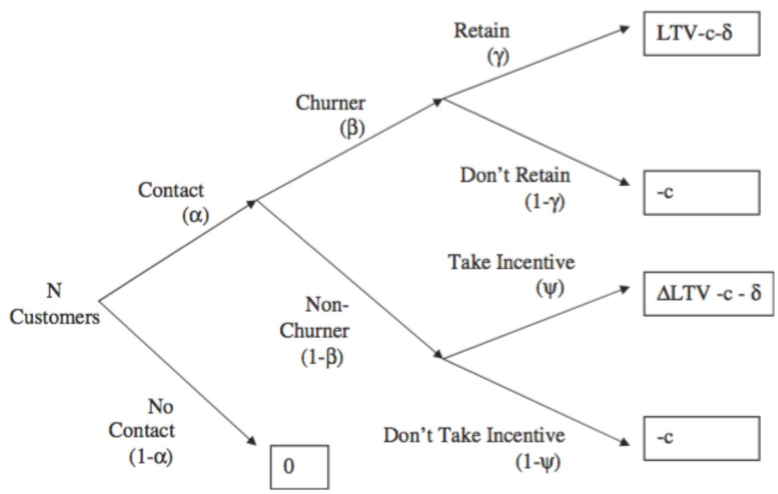


Figure 6: Profitability framework for proactive targeted churn management program, Blattberg et al., 2008.

The presented framework consists of following variables:

N = Total number of customers,

α = The probability a customer is contacted as part of the churn management program,

β = The probability the customer is a churner, given the customer is contacted,

γ = The probability the customer is rescued, given he or she is a churner,

ψ = The probability a non-churner takes the incentive,

Δ = Percentage increase in lifetime value among non-churners who take the incentive,

c = Contact cost,

LTV = Lifetime value of the customer (Blattberg et al., 2008).

One of the first works researching the selection of customers for retention efforts beyond the targeting based on the highest probability to churn is presented by Verbeke et al. (2012). Verbeke et al. (2012) criticize the previous work is evaluating customer churn prediction models based on the top decile lift (top 10% of customers which have the highest probability to churn). The critique is based on the assumption that the profit of this targeting will not be optimal with the choosing the fraction of customers with the highest probability to churn (Verbeke et al., 2012). Therefore, they present the maximum profit criterion which works on the premises that a retention campaign will only be profitable when a company targets a small top-fraction of customers, with high predicted probabilities and with a relatively large sub fraction of true would-be churners (Verbeke et al., 2012). Therefore, the optimal fraction of targeted customers to maximize the returns of a retention campaign will lie within a rather small interval, with the lower bound equal to 0% (Verbeke et al., 2012). The assumptions of this approach are that both the retention rate and average CLV are independent of the included fraction of customers.

Completely different perspective on selecting customers who should be targeted in retention efforts is presented in the study by Guelman et al. (2012). In their two papers studying the uplift models they claim the emphasis of companies should not be on targeting the customers with the highest churn probability, but on the customers who are most likely to positively respond to the retention actions of companies (Guelman et al., 2015). The objective of uplift models is not necessarily the interest in predicting the outcome itself, but in estimating the expected change in the outcome as a result of the action (Guelman et al., 2015). The presented approach is based on the method of random forests, which is tested on a dataset of a major Canadian insurance company and shows incremental gains in retention of customers (Guelman et al., 2015). The downside of this approach is the lack of consideration of profit in targeting of customers, unlike in the approaches of Verbeke et al. (2012) or later studies studies.

One of the most recent studies on the topic of targeting customers for retention was done by Ascarza (2018) where she, just like many other researchers, points out that the previous work for selecting customers for retention efforts is based merely on the risk of churning. Ascarza (2018) bases the research on the fact that previous studies focus on finding which customers are more likely to churn, but no work has investigated whether it is optimal for firms to target those individuals. The suggested approach in customer retention targeting based on this research consists of 3 steps: (1) leveraging the firm's capabilities by running a retention pilot, (2) identifying the observed heterogeneity in the response to the intervention, and based on these steps (3) selecting the target customers on the basis of their sensitivity to the intervention (Ascarza, 2018). The process of selecting customer targets based on their sensitivity to intervention is presented in Figure 7.

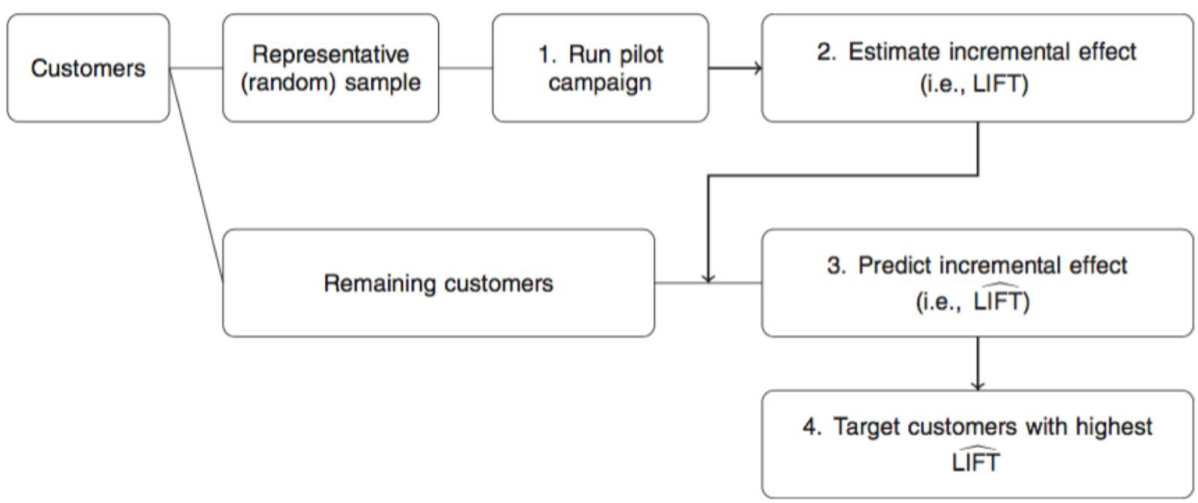


Figure 7: Proactive Churn Management Programs: How to select customer targets on the basis of their sensitivity to intervention, Ascarza, 2018.

Ascarza (2018) suggests that firms should focus their modeling efforts on identifying the observed heterogeneity in response to the intervention and then to target customers on the basis of their sensitivity to the intervention (customers with the highest 'LIFT'), regardless of what is the risk of churning. If a company does not possess information regarding the sensitivity of customers to the company's interventions, the company needs to estimate LIFT by running a small-scale pilot retaining campaign across a representative sample of customers (Ascarza, 2018). Once the heterogeneous treatment effect model is estimated on the pilot sample, the model should be used for the remaining customers to measure their LIFT (Ascarza, 2018). For the last step, the selection of how many customers company should target is presented, the value of LIFT should be also used as a metric for better allocation of resources where the company decides what is the minimum wanted effect of the retention campaign (Ascarza, 2018). The presented proactive churn management of the study is validated by analysing customer behaviour in two field experiments conducted in two different markets (Middle East and North America) and covers two different sectors (professional memberships and telecommunications). By combining the 2 field experiments with Machine Learning techniques, Ascarza (2018) supports assumptions that the presented approach is more effective than targeting customers on the basis of their risk of churning (Ascarza, 2018). However, it is important to note that the findings of this research show the effective approach to reducing customer churn, which is mostly useful in the context where customers pay the same annual fee and the churn is the main differentiator for customer value (Ascarza, 2018). In the business settings, where customer revenue directly depends on consumption (e.g. telecommunications) and customers have different values for a company, the focus on retaining the high-value customers should be taken into consideration (Ascarza, 2018).

A more complex approach to the targeting of customers for retention campaigns is presented by Lemmens et al. (2013, 2017) in their two studies. They criticize the previous approach of targeting only the top decile of customers with the highest churn rates mainly because these approaches do not take into consideration maximizing profits for a company. Lemmens et al. (2013) also points out that the research of Verbeke et al. (2012) solves this issue, however, their research ignores the heterogeneity in the individual targeting opportunity of customers (Lemmens et al., 2013). Therefore, they build on the previous research and propose that the profit of targeting customers depends of four factors: (1) the probability of customer to churn without taking any initiatives from the firm; (2) the value of the customer to the firm; (3) the probability of the customer to respond positively to the retention action and not defect, and (4) the cost of the retention action (Lemmens et al., 2013). This approach of customer heterogeneity in the targeting implies that it is more crucial to accurately predict the churn probability towards some customers than towards others. In particular, the higher the value of a customer with a high churn propensity but a high response probability to a retention intervention, the larger the benefits of targeting this customer. (Lemmens et al., 2013). Furthermore, predictions that overestimate the churn propensity of non-churners, regardless of their value to the firm, are advised to be avoided, especially when the cost of the retention action is large (Lemmens et al., 2013).

In order to design targeted retention programs that aim to maximize profits of retention investments, Lemmens et al. (2013) develop a binary classification method which uses a gain/loss matrix. Afterwards, the customer heterogeneous profit-based loss function and stochastic gradient boosting optimization algorithm are developed to minimize the losses (Lemmens et al., 2013). By using the stochastic gradient boosting, they are able to rank customers by the predicted targeting opportunity. In order to find the optimal size of customers to target, the research is performed on 3 datasets of mature subscribers of a major U.S. wireless carrier where the developed profit-oriented approach is applied (Lemmens et al., 2013). The profit-based step by step analysis is presented in the Figure 8.

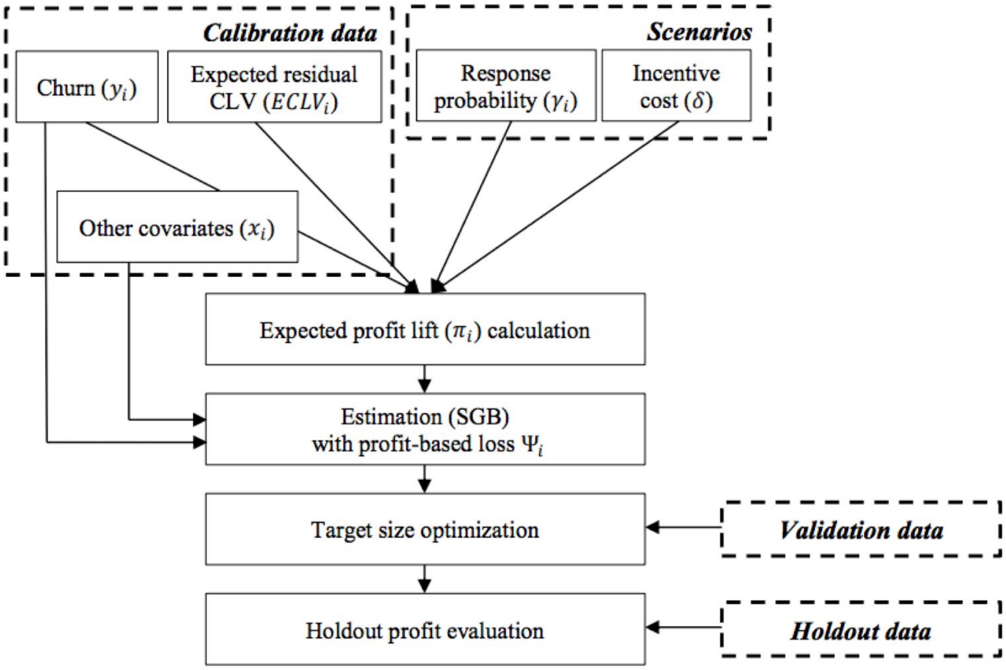


Figure 8: Profit-based analysis step by step, Lemmens et al., 2017.

The results show that profits improve when the target size is optimized rather than determined before and the optimal target size increases with the increasing response rate and decreasing costs of actions (Lemmens et al., 2013). Furthermore, the improvement in financial profits is not correlated with an improvement in the number targeted of churning customers. It is the correct identification of churning customers, that is more profitable for the company (Lemmens et al., 2013). Based on the results of cases, the presented optimal target size and customer ranking lead to improvements in profits compared to the previously used methods (Lemmens et al., 2013).

Drivers of customer churn

Understanding the churn drivers of customers can help firms to get a useful insight for their retention efforts. However, it is important to note that there is a difference between the best predictors of churn and understanding why the customers are churning (Ascarza et al., 2017).

The current research has not necessarily looked at the isolation in causality in churn behaviour, however there are fair number of researches contributing to understanding the causes of churn. With the upcoming interest of customer retention in the 1990s, researchers started to look for the common reasons behind customers' churn decision. DeSouza (1992) suggested companies search for these reasons by directly asking customers, where the information they provide is more specific than information a company can get from a standard market research. DeSouza (1992) presents 6 types of defectors:

1) Price defectors (customers switching to low priced competitor),
2) Product defectors (customers switching to competitor with superior product),
3) Service defectors (customers leaving because of poor service),
4) Market defectors (customers leaving but not to a competitor - e.g. customer moving from the market area),
5) Technological defectors (customers converting to product from outside of the industry),
6) Organizational defectors (customers being lost due to external or internal political considerations) (DeSouza, 1992).

One of the first and most extensive researches of reasons for customers to switch services was conducted by Keaveney (1995). The research is based on the fact that prior work was designed to focus on satisfaction, quality or service encounters as reasons of customer churn, however they do not account for all of the reasons (Keaveney, 1995). Therefore, this study introduces model of customer switching in service industries to help firms to understand the customer defections, which is based on surveying over 500 service customers (Keaveney, 1995). The results of the research identified more than 800 critical traits of 45 different service firms causing customers to switch services, where the reasons of customers switching were classified into 8 categories (Keaveney, 1995). The service switching categories are:

1) Pricing (high prices, price increases, unfair pricing, deceptive pricing),
2) Inconvenience (location/hours, wait for appointment, wait for service),
3) Core service failure (service mistakes, billing errors, service catastrophe),
4) Service encounter failure (uncaring, impolite, unresponsive, unknowledgeable),
5) Response to service failure (negative response, no response, reluctant response)
6) Competition (found better service),
7) Ethical problems (cheat, hard sell, unsafe, conflict of interest),
8) Involuntary switching (customer moved, provider closed) (Keaveney, 1995).

Even though the early research on drivers of customers churning attempted to show some common reasons why customers defect, it is difficult to distinguish how these reasons differ across industries and different business models. Moreover, it is even more challenging to find how these factors influence each churning customer on the individual level.

Most of the studies researching or contributing to the drivers of customers' attrition are conducted in the telecommunication industry environment, due to the highly competitive nature of this industry and the shift of companies to focus their efforts on retaining of customers. Gustafsson et al. (2005) researched the effects of the prominent drivers of retention, which were chosen based on the customer satisfaction and relationship marketing literature. The studied drivers were: (1) customer satisfaction; (2) affective and calculative commitment; and (3) situational and reactional triggers (Gustafsson et al., 2005). The research consisted of qualitative interviews of a large Swedish telecommunication company providing different phone and internet services, and afterwards periodic surveys built on information from the interviews (Gustafsson et al., 2005). The results of the study showed that customer satisfaction and both the affective and the calculative commitment have a consistent effect on decreasing churn (Gustafsson et al., 2005).

The research from Esghi et al. (2007) also investigated the determinants of customers churn in the wireless telecommunication industry. The focus of the study is on the concepts of satisfaction and loyalty and the associations between these concepts (Esghi et al., 2007). The data for this research were collected via phone survey of US telecommunication services consisting of over 2,800 customers, which were analysed by the methods of tetrad model and structural equation model (Esghi et al., 2007). By using these methods, the researchers are able to better understand the determinants of customer loyalty and identify direct and indirect predictors of it, based on which they contribute to the field with several findings (Esghi et al., 2007). Correspondingly to other studies, the findings show that improving customer satisfaction helps to minimize customer churn (Esghi et al., 2007). On the other, the financial incentives lead to misallocation of resources due to the fact that these efforts are mostly aimed at attracting new customers rather than increasing satisfaction of the existing customers (Esghi et al., 2007). Furthermore, the results indicate that there is an importance for service providers focus on new technological solutions in regard to minimize the churn, as well as a strong a association between wireless orientation and propensity to switch providers (Esghi et al., 2007).

Similarly, the research from Seo et al. (2008) focuses on determining the impact of factors of customer satisfaction and switching costs on customer retention behavior (Seo et al., 2008). Furthermore, the research aims to determine whether demographic characteristics of customers have any effect on the customer retention behavior (Seo et al., 2008). To test the hypothesis of the study, the researchers used a binary logistic regression model and two-level hierarchical linear model on the dataset of one of the top ten national wireless service providers in the US, including over 31,000 customers (Seo et al., 2008). Based on the results, the researchers provide several findings. First, there is a strong relationship between switching costs and customer retention behavior, and factors such as reflecting price, handset sophistication, wireless service usage and service plan complexity can increase switching costs (Seo et al., 2008). Furthermore, the results show the importance of technical performance for customer retention, and that even age and gender can indirectly affect customer retention. The differences of these demographic attributes are in regard to the connectivity of wireless service and service plan complexity (Seo et al., 2008).

Yet another approach to the research of major motivators of customer retention was brought by Jeng et al. (2012). The factors and motivators of customer retention of this study were based on a literature review and validation by experts of telecommunication industry and academia (Jeng et al., 2012). The multiple criteria hierarchy framework of these factors is presented in the Figure 9.

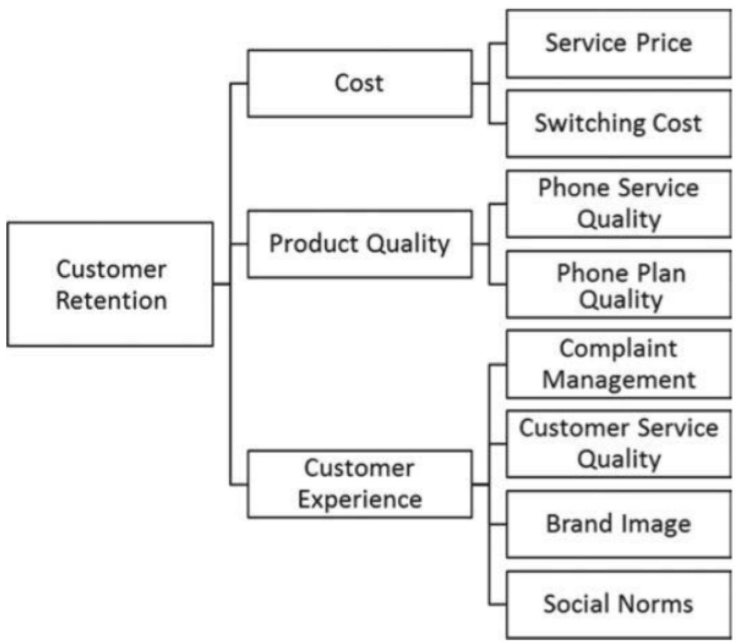


Figure 9: Multiple criteria hierarchy framework, Jeng et al., 2012.

The hierarchy framework was examined in the empirical case of the Canadian mobile telecoms industry, where based on an expert survey, the DEMATEL technique was used for analysis (Jeng et al., 2012). The DEMATEL technique was applied in number of disciplines analyzing complicated social phenomena and is able to show the contextual relations among criteria in the system, showing the strength of the influence of each criterion (Jeng et al., 2012). Based on the empirical research in the Canadian telecommunication industry, the study delivers several findings in regard to motivators of retention. The results showed that cost is the most important motivator in the telecommunication industry and the price is the most identifiable and measurable component of cost (Jeng et al., 2012). The switching costs and quality have moderate importance and are not as important as were in the past, in regard to the customer churn motivators (Jeng et al., 2012). Lastly, the customer experience elements are important but individual elements do not have any special significance (Jeng et al., 2012).

One of the most important contributions to the research of the causes of churn was presented in the study 'Modeling Customer Lifetimes with Multiple Causes of Churn' by Braun et al. (2011). On top of the searching for the reasons behind customers' churn, Braun et al. (2011) points out the need to distinguish between controllable and uncontrollable reasons of churning. By understanding the reasons that are beyond the firm's control, the targeting efforts can be aimed at the customers who are the most likely to react to retention efforts of company (Braun et al., 2011). The research used hierarchical competing-risk framework to model the duration of customer relationship with a company and the reasons why the customer has churned (Braun et al., 2011). Afterwards, Braun et al. (2011) took into account the influence of damper effect, which describes the event when a company takes steps to reduce customer churn by targeting controllable reasons of customer churn but the presence of uncontrollable reasons of churning diminishes company's efforts. The empirical evidence for their approach is presented by US telecommunications firm which provided information regarding their customers to the researchers. By conducting the analysis of the data the research reveals that competing-risk frameworks are required to be used for finding heterogeneous patterns in causes of churn (Braun et al., 2011). Furthermore, understanding how the likelihood of churn is a result of different reasons of churning also provides better understanding how marketing efforts influence customer churn (Braun et al., 2011). Lastly, the researchers suggest that firms shouldn't rely on raw retention metrics since they can mask shifts in the balance between uncontrollable and controllable churn, and misunderstand the effects of company's efforts (Braun et al., 2011).

Nonetheless, Ascarza et al. (2017) in their review of customer retention management point out that identifying specific causes for an individual customer is quite different from identifying general causes in a population. In order to identify the potential causes of churn for an individual customer, the researcher needs to find variables or combinations of variables that are both viable causes and for which the customer exhibits a risky behavior (Ascarza et al., 2017). They suggest that a competing risk hazard model by Putter et al. (2007) in the field of biostatistics may be used to predict which of the possible reasons of churn are most likely to cause churn at any point in time. However, there is one of the more experimental methods where the more reliable methods for identification of individual reasons of customer churn are yet to be discovered (Ascarza et al., 2017).

Predictors of customer churn

As the customer churn prediction models use customer information to present the probability of churn, it is important to understand the relationship between the predictors of customer churn and reasons behind customer churn.

In the afore-mentioned research of Verbeke et al. (2012), where the maximum profit equation criterion is presented, the research extends in conducting an extensive benchmarking experiment. The experiment is evaluating different classification techniques applied on eleven datasets from different telecommunication operators across world (Verbeke et al., 2012). By conducting this experiment, the researchers are able to identify which type of data is the most important and the most suitable for predicting churn (Verbeke et al., 2012). They categorize the selected variables of the best performing techniques and place them into 4 categories: (1) usage; (2) finance; (3) marketing; and (4) socio-demographic variables (Verbeke et al., 2012). The results show that usage variables are the best predictors of churn. The difference in the other 3 types of data in regard to their ability to predict customer churn is minimal, however the results show that complete datasets combined of each type of data have positive impact on predictive performance (Verbeke et al., 2012). Some of the concrete variables which are good predictors of churn are: (1) price and age of the current equipment; (2) the number of contacts between operator and customer; (3) the age of a customer; (4) mean total monthly recurring charge; (5) total number of calls (Verbeke et al., 2012). However, the researchers address the issue in selecting variables because the predictor of churn may not be the reason why churn occurs (e.g. drop in usage may indicate customer is likely to churn, however, sudden peak in usage may also indicate churn, where customer has already decided to attrite and wants to use all the remaining minutes)

(Verbeke et al., 2012). On the other side, the marketing variables are relevant since they can provide actionable information (Verbeke et al., 2012), however, they may not be the best predictors of the churn.

Theoretical conceptualization of designing CRM retention strategies

The purpose of this section is to summarize the findings from the literature review into a conceptual framework (Figure 10.). This framework is illustrated in the context of designing customer retention strategy since it is in line with the main objective of the research. Although, there are already developed frameworks, tackling the topic from different angles, the choice of creating a new framework is to synthesize findings from different sources of literature. Furthermore, in the latter chapters of the thesis, this framework enables the authors to precisely show which steps of the frameworks could be enhanced by Deep Learning Customer Churn Prediction. The foundations for this framework are based on the framework presented by Payne et al. (2006) (Figure 3.) in chapter *Designing customer retention strategies*. Framework presented by Payne et al. (2006) describes strategy development and its implementation with retrospective assessment, which are both aspects of CRM retention strategy that are affected by the Deep Learning customer churn model. However, on top of framework designed by Payne et al. (2006), the new framework includes steps of *degree of individualization of CRM*, *segmentation of customers* and *selection of key customers to target*.

It is important to understand that the several steps of the new framework are put together from different theories, contributing to the process of designing CRM retention in different forms. Certain steps may have direct or indirect influence on other aspects of the framework. However, these effects are not mentioned in the reviewed literature and consequently not illustrated in the framework. Moreover, in different scenarios several steps of the framework may have different order or may be combined into one step. Similarly, certain steps may be divided into more steps, however none of these changes influence the final contribution of the study to show how to implement Deep Learning Customer Churn Prediction model into customer retention strategies.

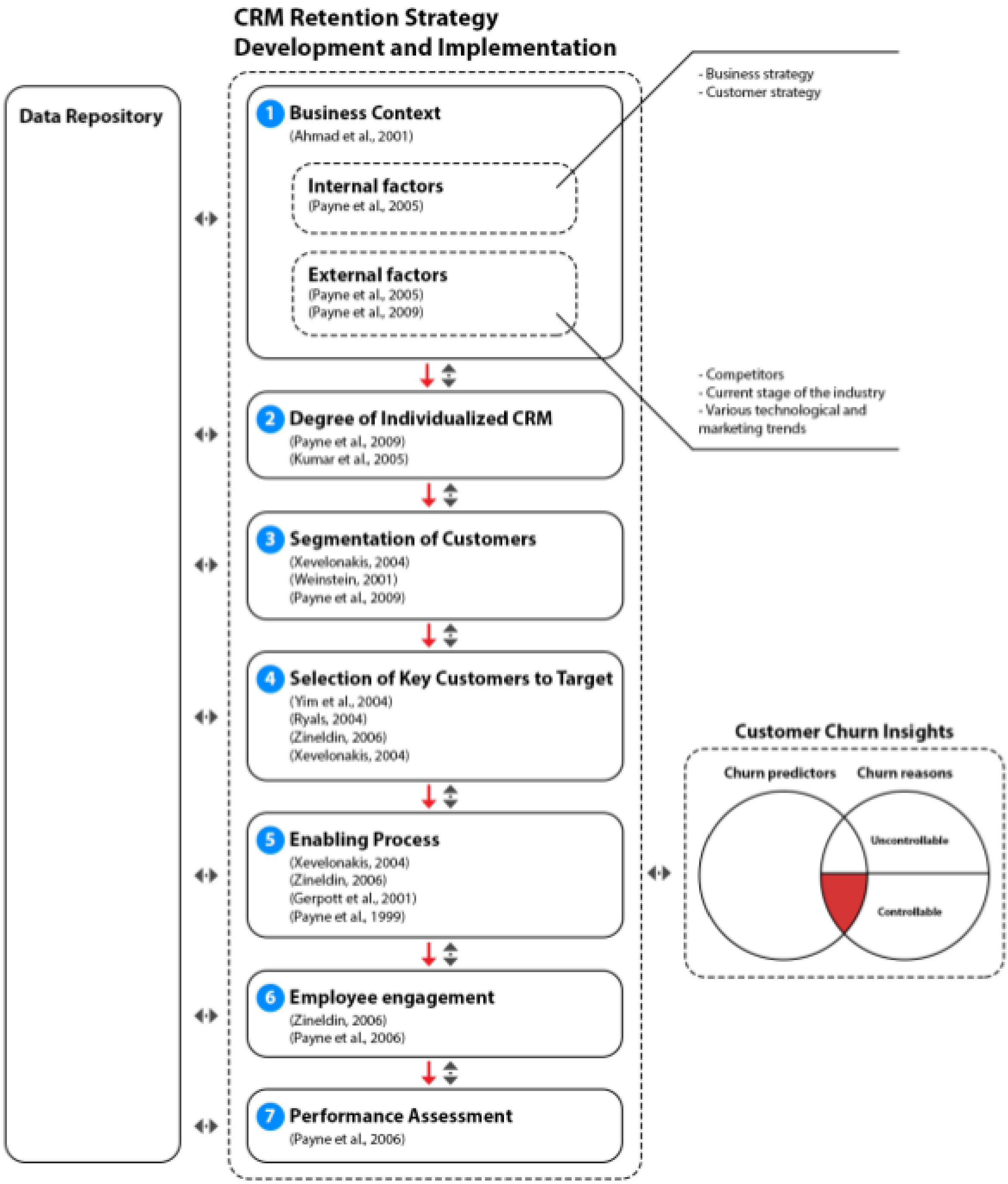


Figure 10: Self-Developed Conceptual Framework of the Process of Development and Implementation of CRM Retention Strategy.

Although, the reviewed literature refers to factors and steps which need to be used in CRM retention strategies, there is a missing argument about what exactly is included in CRM retention strategy, and what is the result of the already set CRM retention strategy. Therefore, the presented framework describes the process of development and Implementation of CRM Retention Strategy, which consist of consequent steps (shown with red arrows). Furthermore, similarly to the framework of Payne et al. (2006), the interaction between the different processes works in both directions where the flow of information influences other steps of CRM strategy (black dotted arrows).

Business Context

Initially, when a retention strategy is being developed, firms need to take into consideration the *business context* consisting of *internal factors* and *external factors* influencing how the strategy should be designed. The *internal factors* an organization has to take into account are the business and customer strategy within the organization, where the process of designing the retention strategy has to be aligned with these two strategies. Furthermore, a firm needs to understand the *external factors* and how they influence the decisions within the company. Amongst these, there are factors such as competitors, the current stage of industry or the various technological and marketing trends within the industry.

Degree of Individualized CRM

The next aspect which company has to decide on when developing CRM retention strategy is the *degree of individualised CRM*. With the shift of CRM and strategy towards the approach of individual-customer-level strategy, companies need to make decision to what extent of individualization the strategy should be designed. The key element of this decision is the completeness of the data a company possesses or aims to possess which enables company to choose the attainable approach. In the presented framework, the data a company possesses is illustrated by the *data repository*. This step is also connected to all other steps due to the fact that decisions in these steps are made with the information acquired from the data. However, it is important to understand that choosing the approach of individualized CRM is not a trivial task, and companies need to consider all the advantages and disadvantages that come with this decision.

Segmentation of customers

Next step in the presented framework is *segmentation of customers*. Several studies in the chapter *Designing customer retention strategies* highlight the importance of segmentation in order to successfully design customer retention strategies. The purpose of segmentation is to build customer clusters and to decide whether the macro, micro or individual-level segmentation is the most appropriate. This step is closely tied to the previous step of choosing the degree of individualised CRM. With a high degree of individualised CRM, customers are approached on an individual level and the one-to-one segmentation is adopted.

Selection of Key customers

The *selection of key customers to target* is strongly recommended step amongst many presented studies in the chapter *Designing customer retention strategies* where some of them suggest selecting customers based on different factors, such as profitability of customers or the probability to churn. This step is further evaluated in the chapter *Selection of key customers to target for retention*. The early work describes the approach of targeting customers with the highest probabilities of churning, where usually the customers in top decile (10%) were targeted. However, the more recent studies show that this approach is not the most beneficial for companies and there are other factors which need to be taken into consideration. There are two main approaches that present improved ways of selecting customers for targeting. Firstly, there is approach of selecting customers based on their response probability to the marketing efforts. The results of this approach show that customer retention may be increased compared to the approaches of selecting customers on their probability of customer churn. The second method does not aim to necessarily minimize churn but to maximize profits from retention campaign. The most complex method presented by Lemmens et al. (2013, 2017) is presented in Figure 8. in the chapter *Selection of key customers to target for retention* and takes into consideration probabilities of customer churn, expected residual CLV, response probability and incentive cost.

This study also highlights that by increasing the accuracy of the customer churn prediction, the profits generated by the retention campaigns increase as well.

Enabling process

A crucial part of customer retention strategy is the *enabling process* where the methods for retaining the selected customers are chosen and implemented. Payne et al. (2006) in their framework describes this step as combination of developing process that extracts and delivers value, multi-channel integration process and incorporating information management processes to acquire valuable information about customers. Other studies highlight information which need to be gained and considered in this step such as measuring customer retention rates, considering customers' perceptions or identifying customer needs. Furthermore, there is an importance of analyzing why customers are leaving as a key aspect of improving customer retention strategy.

Customer Churn insights

The literature review of drivers of customer churn describes the different studies researching reasons behind customers' churn as well as the importance and implications of these reasons. The presented studies by DeSouza (1992) and Keaveney (1995) define several of these defectors across different services and industries, yet the presence and weight of these reasons in the context of subscription-based companies is unclear. However, numerous factors affecting customer retention are present in several studies conducted in the telecommunication industry. Amongst the most mentioned in these studies, there are customer satisfaction, quality and complexity of offered service and price. Nevertheless, apart from understanding what are the reasons behind customer churn, it is important to distinguish between the controllable and uncontrollable reasons in order to improve the retention efforts of a company. The controllable drivers of churn are the ones which can be later targeted by marketing efforts and persuade a churning customer to stay with a company. However, if company aims to to find reasons behind customer churn on an individual level, more experimental methods have to be used since this area needs more research in the future. Finally, in regard to the prediction of churn, it is important to understand that predictors of churn are not necessarily the causes of churn. The combination of variables which perform the best in prediction models are: (1) usage; (2) finance; (3) marketing; and (4) socio-demographic variables, where the ones which are also controllable reasons of customer defection need to be distinguished in order to use them for marketing interventions.

Employee engagement and Performance assessment

The last two steps of the presented framework are *employee engagement* and *performance assessment.* The purpose of employee engagement is to make sure the the developed strategy is delivered appropriately, while establishing the *performance assessment* gathers valuable information for further improvements of the retention strategy.

Deep Learning Customer Churn model

In customer churn prediction a scoring model allows the estimation of a future churn probability for every customer based on the historical knowledge of the customer. In practice these scores can be used to select targeted customers for a retention campaign (De Caigny et al., 2018). In customer churn prediction decision trees and logistic regression are very popular techniques to estimate a churn probability because they combine good predictive performance with good comprehensibility (Verbeke et al., 2012).

Unlike the previous sections in the *Theoretical background* chapter which strive to provide the full picture of a researched topic via systematic literature review, this chapter serves as a foundation for the *Analysis* chapter. It provides a brief description of Deep Learning, types of Neural Networks, an overview of related work, describes interpretability of Deep Learning networks and lastly lists two requisite steps with regard to data preprocessing in order to build a Deep Learning architecture. As described in the chapter *Literature Review Process,* this choice was made due to the practicality of the topics discussed on the following pages. Furthermore, as far as putting the networks into practice, this field is still fairly young and all of the research is done in a case-specific setting. Nonetheless, the papers discussed in this chapter of the *Theoretical background* provide the researchers with the most contemporary knowledge of the problem.

Lasly, while it is not so challenging to use Neural Networks for a case problem, the mathematical explanations and equations behind these networks are extremely difficult to fathom and describe in a marketing thesis. It is for this reason that the authors keep the descriptions of the Neural Networks fairly brief and rather highlight their utilization.

Deep Learning Neural Networks

The aim of Artificial Neural Networks (ANN) is to reproduce the inner workings of the human brain. These algorithms employ interconnected elements, or neurons, working together as one system to solve a specific problem. The objective is to find structures, patterns and connections within a training dataset with their ability to learn automatically from available data in order to provide a means for predictions (DeMuro, 2018). The advances in the field of Neural Networks and recent increases in computing performance have allowed for the development of large-scale Neural Networks. This leads to the networks being able to decompose the complexities within the given data by generating abstract data features in an unsupervised manner in each of their hidden layers (Bengio et al., 2013; Bengio, 2009).

The development gave new life to predictive modelling on high-dimensional datasets with very noisy data as the unsupervised abstract features were able to capture the most important variances within the data and thus ignore any variance that did not affect the result variable (Deng et al., 2013). This inherent ability has made Deep Neural Networks (DNNs) excellent tools in pattern recognition. Since churn prediction is the analysis of user behavioural patterns, the application of DNNs in this domain could definitely be beneficial not only in terms of prediction accuracies but also in eliminating manual feature engineering as a required step (Spanoudes et al., 2017).

Types of Deep Learning Neural Networks

The number of Neural Network architecture rises exponentially and only in 2016 the Neural Networks summary created by Fjodor van Veen from The Asimov Institute comprised 27 types of networks. For this reason, this chapter provides three examples which are the most prominent within the research of customer churn. These are Fully connected Neural Network, Recurrent Neural Network and Convolutional Neural Network (Tch, 2017).

A Fully connected Neural Network comprises one or a number of fully connected layers is shown in Figure 11. A multi-layer Fully connected Neural Network is then presented in Figure 12. In such a layer, each output is therefore dependent on each input (Ramsundar, 2018).

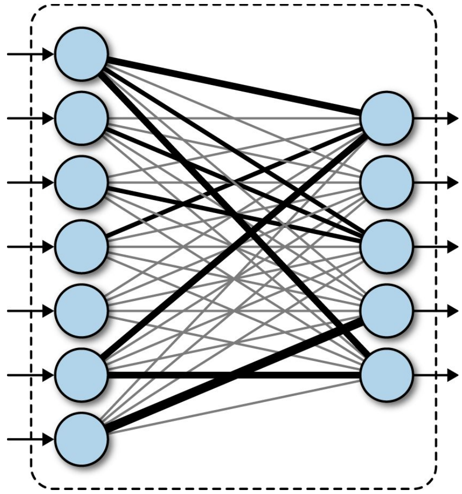


Figure 11: Visual representation of a fully connected layer, Ramsundar, 2018

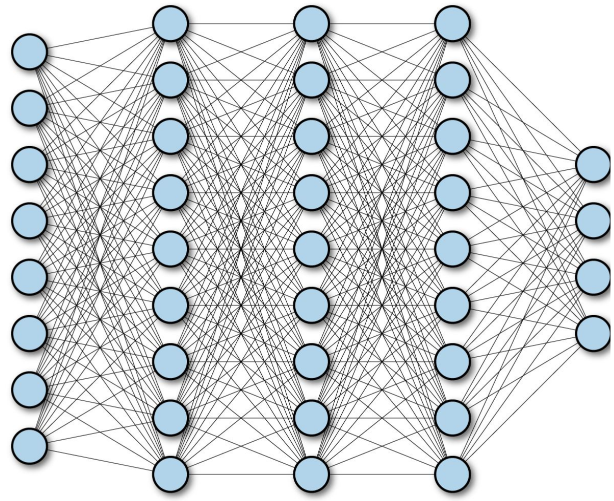


Figure 12: Visual representation of a fully connected Neural Network, Ramsundar, 2018

The advantage of Fully connected Neural Networks is that they are flexible in terms of the input. On the other hand, this generality brings the issue of lower performance in specific-input cases (Ramsundar, 2018). Usually, this type of network is used in structured-input cases as well as classification and regression problems (Brownlee, 2018).

Recurrent Neural Networks are specific by saving the output of a layer and parsing it back to the input which improves the final outcome of the layer. This way, each neuron retains some information it learned in the previous iteration. Long Short-Term Memory (LSTM) is then an extension of RNN which enlarge their memory capacities by remembering the inputs for a longer period of time (Mehta, 2019; Donges, 2018). The usage of Recurrent Neural Networks primarily revolves around text-to-speech conversion (Maladkar, 2019). LSTM, however, has a great capacity to deal with time-series datasets (Malik, 2018).

Lastly, a Convolutional Neural Network is a set of one or multiple Convolutional layers and is mostly used in Image or Video recognition. Their ability to work well with spatial information, more specifically to learn shapes, positions and scales is a perfect prerequisite for such an utilization (Brownlee, 2018).

Related work to the KKBox Deep Learning architecture

Churn analysis using multiple data sources

B2B cloud computing service Azure powered by Microsoft employs Deep Learning Customer Churn model based on two streams of client information - billing information and usage data throughout a time period of 8 weeks. The former includes subscription type, length of the subscription or segmentation information. As for the time-series data, Microsoft have at their disposal more than 20 variables including overall consumption, usage of particular services or free trial usage (Zhu et al., 2017).

The data scientists at Microsoft opted for Deep Learning architecture thanks to its ability to yield far more satisfactory results when it comes to the accuracy of the model then with the basic linear models. In order to cope with the accuracy vs interpretability phenomenon, the cloud computing service uses previously mentioned LIME to provide comprehensible insights into the predictions. The architecture which Microsoft had patented looks like as follows (Zhu et al., 2017; Cornelisse, 2018):

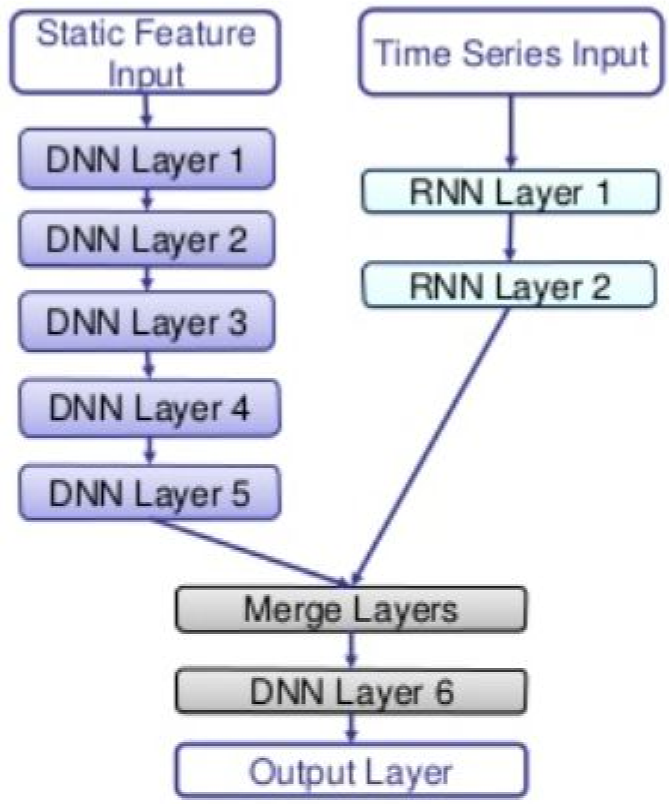


Figure 13: Visual representation of Microsoft Azure's architecture, Zhu et al., 2017

Microsoft describes the model as "hybrid" conjuction of Deep Neural Network branch (DNN) and Recurrent Neural Network branch (RNN). The former uses basic algorithm Multilayer Perceptor (MLP) to process the static data while simultaneously finding patterns in the time series data using Long Short-Term Memory (LSTM) networks. The input for the latter, right branch, is in the shape of 18*56, meaning 18 time-series features across 56 days. The outputs of each branch are later merged into a single output. Lastly, the scientists use LIME for a customer on an individual level to detect which features contributed to the particular classification. This creates the basis for customer intervention (Zhu et al., 2017; Cornelisse, 2018).

Churn analysis using image recognition

Just as in the case of Microsoft Azure, the researchers from King Mongkut's University of Technology in Thailand had at their disposal a dataset representing consumer behavior. The collaboration took place in 2016 between the University and communications conglomerate True

Company. The subject of the churn analysis were therefore the prepaid customers of True Company's telecommunications services. These data therefore included ten behavioral traits such as data usage, top up amount, top up frequency or voice calls (Wangperawong et al., 2016).

The data scientists utilized the exceptional capability of Deep Learning related to image classification in comparison to the regular Fully connected network, and recreated user behavior as in form of a heatmap, see Figure 14. (Wangperawong et al., 2016).

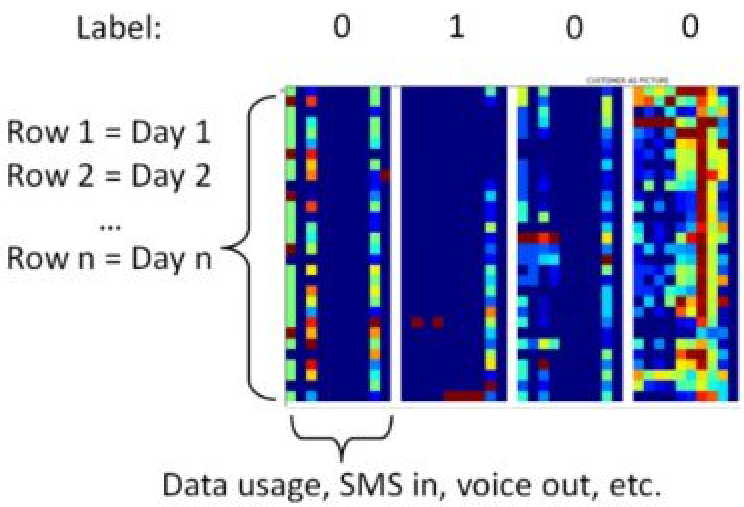


Figure 14: Heatmaps representing customer behavior of 4 customers - non-churner, churner, non-churner, non-churner (left to right), Wangperawong et al., 2016

Each heatmap, representing each user's behavior, includes the 10 behavioral characteristics (on the x-axis) within the timespan of 30 days (y-axis) (Wangperawong et al., 2016). The input is then parsed into the Deep Learning model for image recognition as shown in Figure 15.

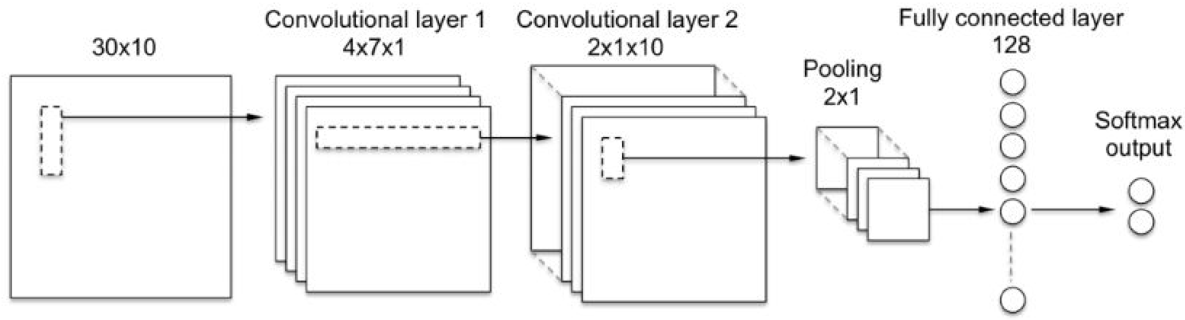


Figure 15: Convolutional Neural Network used by Wangperawong et al. to predict customer churn, Wangperawong et al., 2016

The architecture comprises two convolutional layers, one pooling layer, a fully connected layer and softmax output. The first layer consists of 4 filters with the shape of 7x1 which aims to find weekly traits (7 days) in each individual usage behavior (1 column). The output is then forwarded to the second convolution layer with the shape of 1x10 which tries to find patterns in each timestamp (1 day) across all the usage behavior (10 columns). The consequent pooling layer is applied to secure the optimal number of parameters, to reduce the training time and to manage overfitting (Wangperawong et al., 2016; Cornelisse, 2018). The fully connected layer flattens the output and lays the foundation for the binary classification rendered by the softmax output (Wangperawong et al., 2016).

Interpretability of Deep Neural Networks

While there has been an enormous demand for looking under the hood of Deep Learning models, it is noteworthy to stress out that the field is still fairly young and unexplored. It is, among others, for this reason that the more easily interpretable linear models and decision trees still prevail in usage in certain areas (Montavon et al., 2018). This phenomenon is called accuracy vs interpretability problem and describes the situation where the data scientists face a choice whether to settle for a less complex worse-performing model due to the need for understanding the results. Or if they better use a better model at the expense of full comprehensibility (Goodrum, 2016).

As a result, many ways are being introduced as to how to interpret even the models striving for higher accuracy. There are two branches of methods approaching the interpretability of Deep Neural Network models. As saliency methods provide an insight as to what happens within the Neural Network itself, they are rather useful for the data scientists trying to determine the best approach to build a Neural Network architecture. Feature attribution methods, on the other hand, are suited for linking the input variables to the output and answering questions such as "What variables had the greatest impact when predicting the results? (Perez Denadai, 2018)" Lundberg et al. (2017) present explanation models, namely LIME, DeepLIFT and Layer-Wise Relevance Propagation as methods belonging to the latter interpretability family (Lundberg et al., 2017).

LIME, an abbreviation for Local Interpretable Model-agnostic Explanations, is a model introduced in a paper "Why should I trust you?" by Ribeiro et al. (2016) The authors describe the model as such: "a modular and extensible approach to faithfully explain the predictions of any model in an interpretable manner," (Lundberg et al., 2017). The method gives an insight into the features that were used for the prediction on an individual level, meaning locally for each output. As a result, a set of variables is given on top of each prediction with either positive or negative impact on the actual prediction (Lundberg et al., 2017; Ribeiro et al., 2016).

Unlike LIME which adapts to any model, Deep Learning Important FeaTures or DeepLIFT is a method tailor made for Deep Learning architectures. The model introduced in 2017 uses difference-from-reference approach with "importance scores based on explaining the difference of the output from some 'reference' output in terms of differences of the inputs from their 'reference' inputs (Shrikumar et al., 2017)." Similarly to LIME, DeepLIFT provides both positive and negative contributions to the prediction (Shrikumar et al., 2017).

Layer-wise Relevance Propagation (LRP) is once again a method that examines and assesses the relevance of separate inputs via backward pass through the Neural Network architecture as shown in Figure 16. In this particular case, the aim of the image recognition architecture is to find the most relevant pixels in the image (upper chart). Sequently, the backward pass procedure starts with the relevance summed up in the output layer and iterates back through the Deep Neural Network (lower chart). As a result, the fragmented predictions are presented in the same manner as the input information (Montavon et al., 2017; Shiebler, 2017).

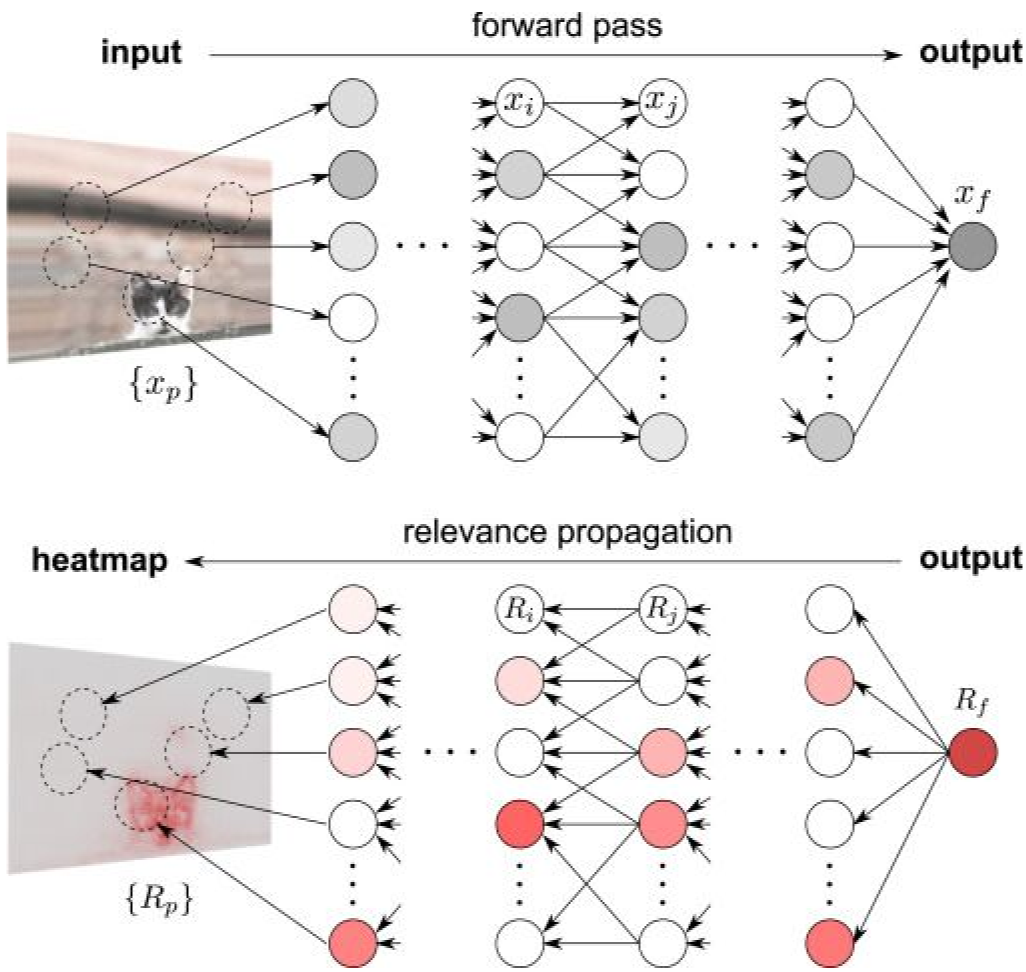


Figure 16: Layer-wise Relevance Propagation backward pass, Montavon et al., 2017

Imbalanced class distribution

Even though customer churn is of great importance to companies, in the usual sample of the entire customer base it occurs quite rarely. This creates class imbalance in the customer data where the non-churning customers take up the majority of the dataset while the actual churners are heavily under-distributed. Imbalanced datasets is one of the most challenging data science problems and becomes the focus on thorough research. The substantial bias towards the larger class (non-churn) when training the algorithm heavily decreases the accuracy of the model. The two most common approaches to the issue are either over-sampling or under-sampling. In the former case, a model is used to equalize the class distribution by creating artificial copies of the under-sampled class. The latter method randomly excludes samples of the over-distributed class (Gui, 2017).

Train and Test split

The Train and Test step is a crucial step in the Machine and Deep Learning process. These techniques of Artificial Intelligence work in a way that they are given input together with a known class (in the case of customer churn, these are churning and non-churning customer) and based on the traits and patterns in the input dataset, they learn which of these belong to which class. Therefore, it is imperative to provide the Deep Learning architecture with two subsets of the entire dataset (see Figure 17.) (Bronshtein, 2017).

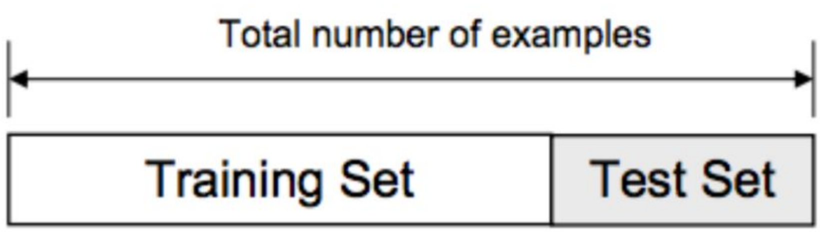


Figure 17: Splitting dataset into training and testing subsets, Bronshtein, 2017

The Training subset usually takes up 80% of the full dataset and is used by the model to find the traits belonging to either class. The remaining 20% is called Testing dataset and is used for validation of the predictions (Bronshtein, 2017).

Analysis

The following analysis comprises two interconnected parts. In the first part, the researchers build the Deep Learning architecture and examine the viability of the combination of the two approaches mentioned in *Related work to the KKBox Deep Learning architecture* in a real-market setting.

In order to construct the Deep Learning Customer Churn Prediction model in a real-market environment, the researchers opted for data available online from a Taiwanese music streaming service KKBox. The datasets were part of a case initiated by the KKBox company itself via site Kaggle which is online community specializing in collecting projects, cases and best practices for data science and Machine Learning. The objective of the competition was to identify whom of the KKBox customers will defect during the next period based on the data regarding the behavior on the service as well as transactional and demographic data. Nevertheless, the data remained available for learning purposes even after the competition has been closed at the end of 2017 (Kaggle.com, 2017).

Furthermore, the researchers attempt to derive additional marketing insights in the form of churn predictors. The first part thus spans across the following chapters:

- *Data exploration and preprocessing*
- *Customer churn prediction model*

The second part of the Analysis is then a combination of the yielded results from the first part with the outcome of the Theoretical background in the form of the Conceptual framework. This is precisely discussed in the chapters:

- *Synthesis of the Conceptual Framework and customer churn prediction model*
- *Marketing application*

Data exploration and preprocessing

Data exploration is the preliminary process of understanding the collected raw pieces of information. It can unveil hidden relationships and provide a company with valuable insights even without any further analysis. Consequently, it helps the researchers better understand the data at their disposal and later serves as a proxy to make more informed choices about data processing and analysis. On the other hand, data exploration itself is the most time-consuming part of the entire data analysis with around 70% of the assignment time to be devoted to it (Ray, 2016).

The following data preprocessing is a technique which aims to transform the raw data into a more understandable form of information. Mohit Sharma (2018) explains the data preprocessing specifically for Machine Learning in six steps (Sharma, 2018):

1. Importing specific libraries for data preprocessing
2. Importing the raw data
3. Exploring missing values
4. Handling categorical variables
5. Splitting the datasets into Train and Test
6. Feature scaling

Firstly, importing computing, statistical and mathematical libraries within the programming language is essential as it provides a far easier approach to the following steps in the preprocessing framework. These libraries also help with fetching and importing the raw datasets (Sharma, 2018). As stated before, the researchers used programming language Python and several preprocessing packages within the language. As for fetching the data, the researchers obtained the data directly from the Kaggle database via Kaggle API. The data comprises four datasets in total - demographic data about the customers in the Members dataset, transactional data in the Transactions dataset and behavioral data with regard to the daily usage of the service in User logs dataset. Apart from the features which will help the Deep Learning algorithm classify the customers, the fourth dataset consists of a label for each user - churner or not churner.

After the raw data have been properly explored, the first step in the actual data preprocessing is managing missing values. There are three ways to handle this issue. If a feature misses more than 75% of the values, it is advisable to remove the entire column. With less than three quarters of the feature values missing, it is fitting to replace the missing values either with mode, median or mean of the present values. At the same time, missing values could be also imputed by a predictive model, namely either decision tree or linear regression (Joseph, 2016; Sharma, 2018). As none of the features was missing more than three quarters of its values, the researchers employed mean values as well as decision tree to impute missing values in various features in the Members and Transactions datasets. Furthermore, an issue-specific preprocessing step with regard to replacing values in a dataset which is missing in the framework by Sharma is detecting and replacing the outlying values. With all three datasets containing outlying values, the detailed process of handling outliers is described in each respective dataset chapter.

Machine Learning in general works with numerical values therefore it is imperative to transform the categorical variables in form of names or ID's into their numerical representation in form of dummy variables. Dummy variables create as many columns as there are values in a feature and take form of 1 in the column which represents the value in question and fills the remaining columns with 0s (Sharma, 2018). Again, the specific steps with regard to dummy variables are described in the respective preprocessing chapters.

As mentioned before, Machine Learning algorithms take in a part of the entire dataset, learn from that by finding patterns and important features and validates the accuracy by the remainder of the dataset. Usually, the entire dataset is split by the ratio of 80:20 where the majority is used for the former, hence the name training dataset, while the remaining 20% is used for validation - the testing dataset (Sharma, 2018). This step is thoroughly described in the chapter *Train and Test split*.

As the last step in Sharma's framework, feature scaling brings all the features, which may differ heavily in figures, to the same scale. This helps to restore the dominance in feature importance for variables with higher nominal figures - for example salary (in tens of thousands) at the expense of age (in units) (Sharma, 2018). Feature scaling is recommended in any case of Machine Learning and can be most frequently achieved one of the following methods: standardization, mean normalization, min-max scaling and unit vector. The researchers used min-max scaling for all the continuous variables in the Members, Transactions and User logs datasets (Kathuria, 2019). The equation of min-max scaler is shown in Figure 18. with its representation in Python language in Figure 19.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Figure 18: Min-max scaling equation, Kathuria, 2019

```
def normalizer(x,min_value,max_value):
    return (x-(min_value))/(max_value-(min_value))
```

Figure 19: Code snippet: Python function for min-max scaling

Furthermore, there are more issue-specific steps needed which are missing in the Sharma's data preprocessing framework. These include padding followed by transformation into a heatmap in the case of the User logs dataset. Padding is the process of bringing arrays of variables to the same length (Brownlee, 2017). As the number of timestamps vary for each user within the time period of 59 days, this is an imperative measure in order to accomplish rendering customer behavior as a heatmap. Again, both of these steps will be thoroughly described in the *User logs* chapter. Another issue-specific preprocessing step includes feature engineering. Feature engineering is the process of extracting additional information from the raw data and hand in hand with Machine Learning modeling (described in the chapter *Analysis*) represents a technique how to maximize the accuracy of the prediction outcome (Koehrsen, 2018). The researchers employed this method while aggregating the data in the Transactions dataset.

Overall, the researchers have at their disposal 3,763,505,439 data points spread across the three datasets. However, the datasets needed to be cleared off of the users who are not present across all three datasets since the three-branch Deep Learning model described in the Analysis chapter requires data from each source for each user. Consequently, all the datasets needed to be sorted in the same order to be simultaneously parsed into the Deep Learning model. Moreover, due to the multitude of data, the researchers limited the period for Deep Learning classification to only the first two months of 2017, specifically 59 days between the 1st of January and the 28th of February. This way, the analysis employs data about 753,856 users.

Members

The members dataset contains 6,769,473 rows, meaning information about 6,769,473 unique users. Each user row includes the following 6 features (Kaggle.com, 2017) present in Table 3.

Variable	**Variable type**	**Description**
msno	Categorical	User id
city	Categorical	City of the user
bd	Ordinal	User age
gender	Dichotomous	Use gender
registered_via	Categorical	Registration method ID
registration_init_time	Ordinal	Registration date

Table 3: List of variables in the Members dataset including description

City

The geography distribution of KKBox users is presented in Figure 20. with the absolute number of users on the x-axis and the city on y-axis. Unfortunately, the cities are represented only by IDs, presumably due to confidentiality reasons.

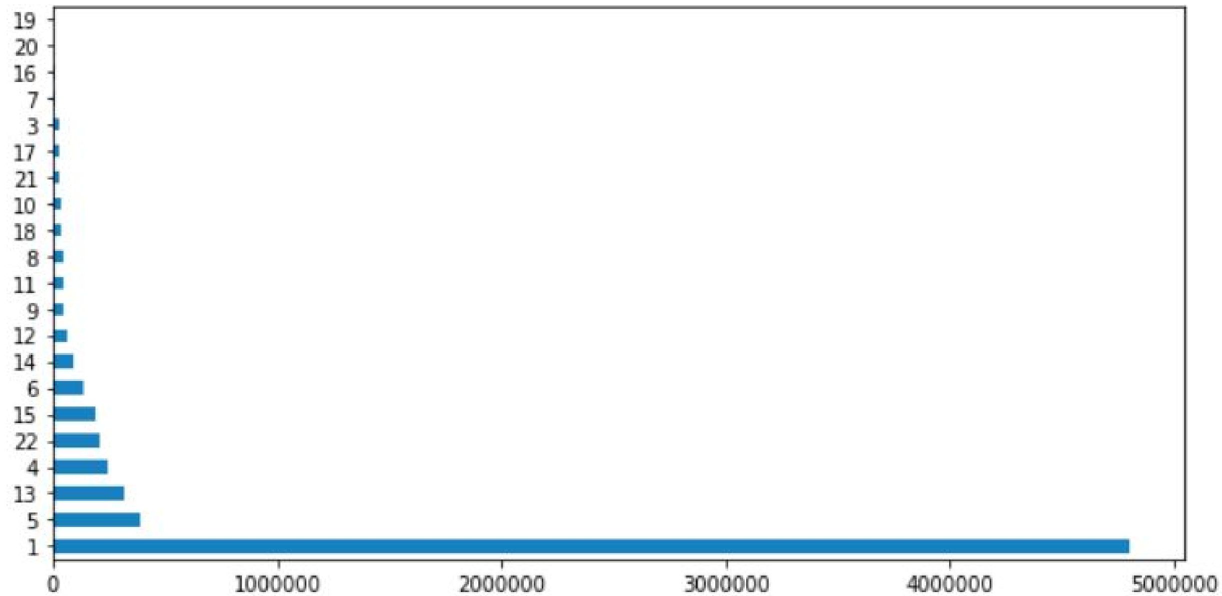


Figure 20: Geographical distribution of KKBox users (city names decoded under city ID's)

Age

The column regarding the users' age contained both missing and outlier figures. The column contains 4,540,215 missing values out of 6,769,473 meaning two thirds of the column are empty. The few outliers vary from values -7168 to 2016.

The following Figure 21. shows the logarithmic scale of the values across the age column. Important fact to keep in mind is that the chart represents a logarithmic scale, not an actual distribution of values. This measure was taken in order to diminish the skewness towards high-count values in form of missing values and proper age representing 67,06% and 32,78% of the dataset, respectively (Robbins, 2012).

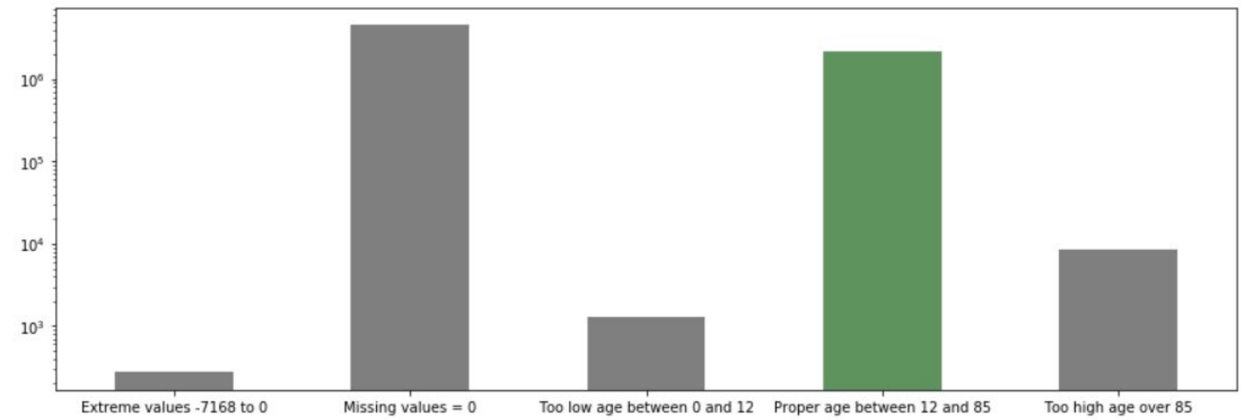


Figure 21: Age distribution of KKBox users

Gender

The column regarding gender of the KKBox users was the column with the highest percentage of missing values. Out of 6,769,473 unique users, it contained 4,429,505 rows where the user was hesitant to state their gender. While the missing values make up for about 65% of the dataset, according to data preprocessing guidelines presented by Mohit Sharma, the column does not have to be terminated unless there are more than 75% of the values missing (Sharma, 2018). The following Figure 22 shows the distribution of stated genders.

```
missing    4429505
male       1195355
female     1144613
Name: gender, dtype: int64
```

Figure 22: Gender distribution of KKBox users before preprocessing

Registered via

There are 18 registration methods, however, the obtained dataset contains only IDs of the methods and there is no additional explanation provided. Therefore, the researchers do not get to see what the actual method is and cannot derive any further implications.

The distribution of the methods is as follows with the 4 most prevalent methods making up for 99,34% of the dataset.

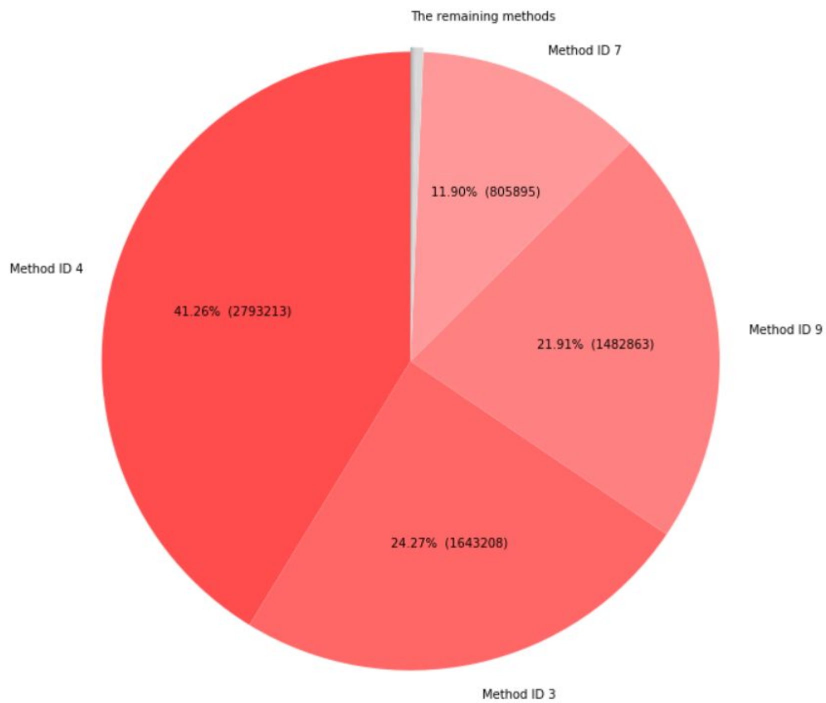


Figure 23: KKBox user-base split in terms of registration method

Registration time

The registration timeline stretching over 13 years from the 26th of March 2004 until 29th of April 2017 looks like as follows:

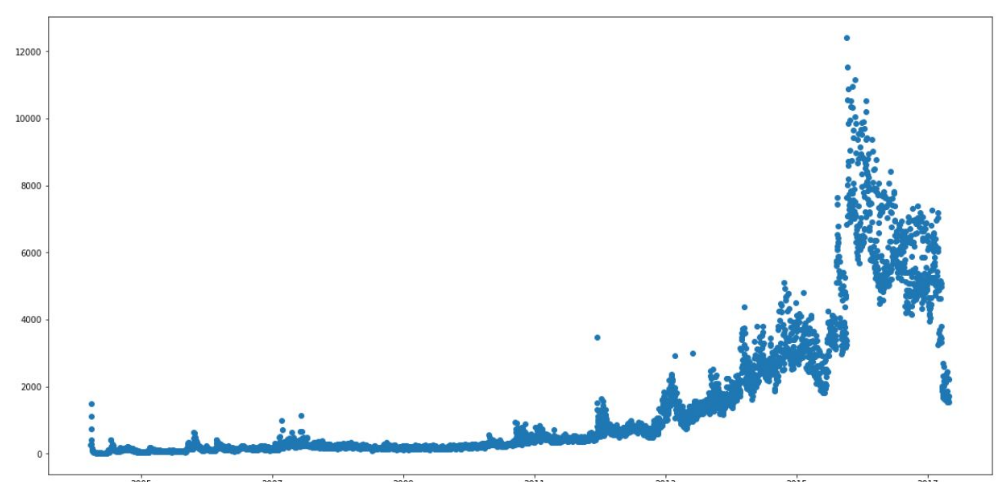


Figure 24: Number of registered KKBox users per day between 26th of March 2004 and 29th of April 2017

The peak takes place on the 9th of October 2015 with 12,413 registered users in one day. On the other hand, the lowest point took place early on in the KKBox history on the 25th of May 2004 where only a single user signed up for the service.

Deep Learning Input preparation

The Members dataset preprocessing consisted of the following steps: detecting and replacing outlying values, imputing missing values, dummy variables, feature engineering and feature scaling. The entire preprocessing code (with the exception of imputing gender missing values shown later in Figure 28.) is present in the following Figure 25:

```python
def preprocessing(input_df,filename):
    input_df = input_df.sort_values(by=["msno"])
    input_df = input_df.reset_index(drop=True)

    input_df["tenure_length"] = [(datetime.strptime("20170228","%Y%m%d")
                                 - datetime.strptime(str(i),"%Y%m%d")).days for i in input_df["registration_init_time"]]

    dummies_city = pd.get_dummies(input_df["city"]).add_prefix("city_")
    dummies_gender = pd.get_dummies(input_df["gender"]).add_prefix("gender_")
    dummies_registration_method = pd.get_dummies(input_df["registered_via"]).add_prefix("rmethod_")

    members_preprocessed = pd.concat([input_df["bd"],
                                      input_df["tenure_length"],
                                      dummies_city,
                                      dummies_gender,
                                      dummies_registration_method], axis=1)

    for feature_name in members_preprocessed.columns:
        members_preprocessed[feature_name] = normalizer(members_preprocessed[feature_name],
                                                        min(members_preprocessed[feature_name]),
                                                        max(members_preprocessed[feature_name]))

    with h5py.File(filename, 'w') as hf:
        hf.create_dataset('members', shape=(members_preprocessed.shape),
                          maxshape=(members_preprocessed.shape), dtype = 'float32')

    members_preprocessed.to_hdf(filename, 'members', mode='w')
```

Figure 25: Code snippet: Preprocessing of the Members dataset (excluding imputation of gender missing values)

Firstly, as illustrated in the previous chapter regarding data exploration, the values in the age feature vary between -7168 to 2016. A few records are between the figure 1937 and 1970 which the researchers assessed as the users' confusion between the age and date of birth when signing up. These values were therefore replaced by a corresponding age. Consequently, both missing values and outliers were replaced by the weighted mean value of the age values between 13 and

85 which the researchers evaluated as a truthful age range for using KKBox's services. The mean value being 29.41 thus replaced 67,2% of the dataset.

The following chart represents the preprocessed age distribution. Once again, it is noteworthy that the chart features a logarithmic scale, not a linear one due to the skewness toward the imputed mean value making up for two thirds of the dataset (Robbins, 2012).

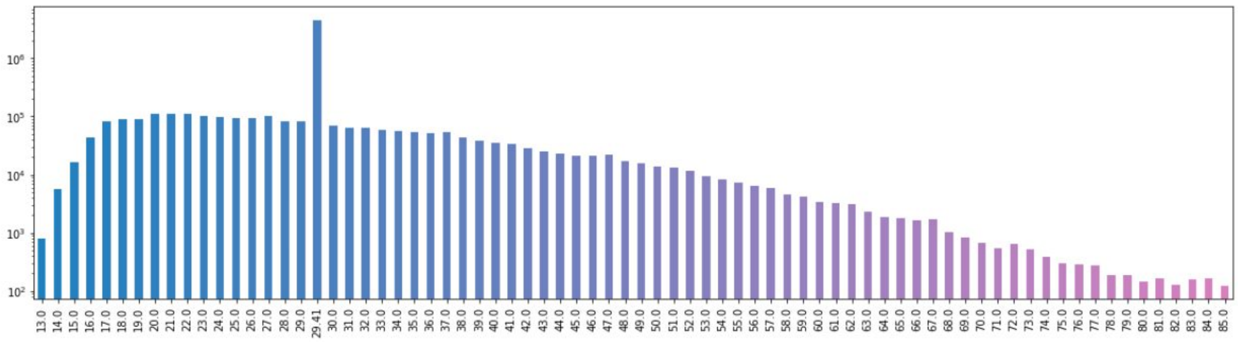


Figure 26: Logarithmic scale of the age distribution of KKBox user-base after handling missing and outlying values

In the case of the Gender feature, the researchers opted to impute the missing values. In this case, two options are available as to how to replace the missing values. One approach is to impute the values by average, which given the size of the dataset and the percentage of missing values is undesirably straightforward. The second option is to employ decision tree to use predictions based on other variables (Joseph, 2016). The other variables available for predicting the gender were age, city and registration method. Since age or city should not have any correlation to the gender, the researchers created a simple decision tree based on the method the user chose to register.

A crosstab between gender and the registration method is available in Figure 27. For each registration method it exposes the percentage of the gender entirety using respective method to sign up for the service. It can be seen that the methods with id 3, 9 and 17 are prevalent among females while the rest among male.

```
gender             female      male max_value
registered_via
1                0.000001  0.000003      male
2                0.000047  0.000098      male
3                0.521496  0.514121    female
4                0.105788  0.113109      male
5                0.000044  0.000073      male
6                0.000014  0.000035      male
7                0.045786  0.052241      male
8                0.000115  0.000211      male
9                0.324414  0.317382    female
11               0.002034  0.002411      male
13               0.000157  0.000196      male
14               0.000069  0.000081      male
17               0.000009  0.000006    female
19               0.000026  0.000034      male
```

Figure 27: Registration method per gender, basis for the decision tree for gender imputation

Each missing figure was distributed between males and females based on the prevalent registration method within either gender. The code snippet showing the decision tree is shown in Figure 28.

```
for index,row in members.iterrows():
    if row["gender"] == "missing":
        if row["registered_via"] in [3,9,17]:
            members.at[index,"gender"] = "female"
        else:
            members.at[index,"gender"] = "male"
```

Figure 28: Decision tree used to impute missing gender values

The loop goes through each row of the dataset and detects the missing values. If the row's (user's) registration method ID is 3, 9 or 17, it replaces the missing gender value with female. Otherwise, male gender is inputed.

Figure 29. then examines the Gender column after preprocessing by the above-described decision tree.

```
male      4459493
female    2309980
Name: gender, dtype: int64
```

Figure 29: Gender distribution after decision tree imputation

The categorical variables, namely City, Gender and Registration method, were represented by dummy variables where each value (City ID, Gender and Registration method ID) takes form of a unique column. As for the Registration time, the researchers used the values to calculate the tenure length of each user. Lastly, each column in the dataset was brought to the same scale between 0 a 1 via min-max scaler. As a result, the input shape of the Members dataset was 753856x30, meaning 30 variables about 753,856 unique users.

Transactions

The Transactions dataset represents payment data of 2,363,626 unique users between the period of 1st of January 2015 and 28th of February 2017. In total, the dataset comprises 21,547,746 rows, each consisting of 9 features, as further described in Table 4. (Kaggle.com, 2017).

Variable	Variable type	Description
msno	Categorical	User id
payment_method_id	Categorical	Payment method
payment_plan_days	Ordinal	Length of membership (days)
plan_list_price	Ordinal	Price of membership (New Taiwan Dollar)
actual_amount_paid	Ordinal	Actual amount paid (New Taiwan Dollar)
is_auto_renew	Dichotomous	Indicates whether the membership is auto renewal
transaction_date	Ordinal	Date of the transaction
membership_expire_date	Ordinal	Date of the membership expiration
is_cancel	Dichotomous	Indicates if the user canceled the membership during the transaction

Table 4: List of variables in the Transactions dataset, including description

Payment method and Payment plan

The following chart describes the distribution of the payment methods among the KKBox's users. Despite the fact that the values are recorded under ID and no further description is included, it is clear that there is one prevailing payment method making up for 53,5% of the dataset.

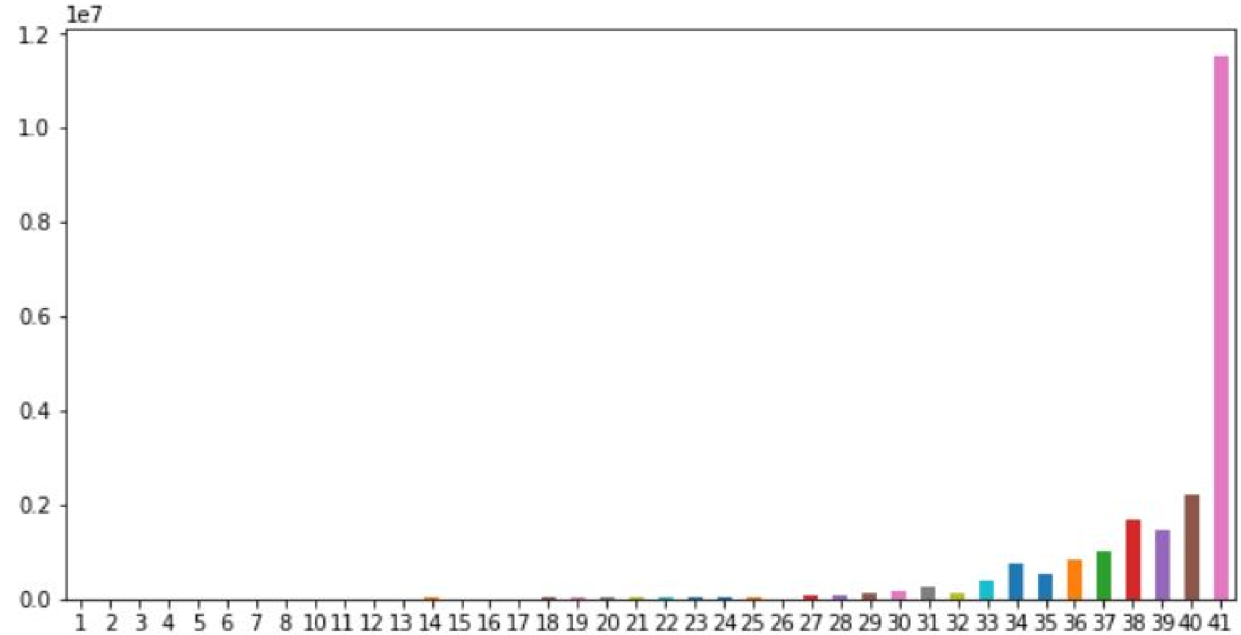


Figure 30: Distribution of payment methods among KKBox users (methods decoded by method ID's)

Payment plan, on the other hand, is the amount of days a user prepays for (subscribes) each time; in other words, the length of each subscription. It can be seen that the most common 30-day term is most frequent, however, over the entire user base, users choose to prepay from 1 day to as many as 450 days in advance. The distribution of the length of the subscription is as follows:

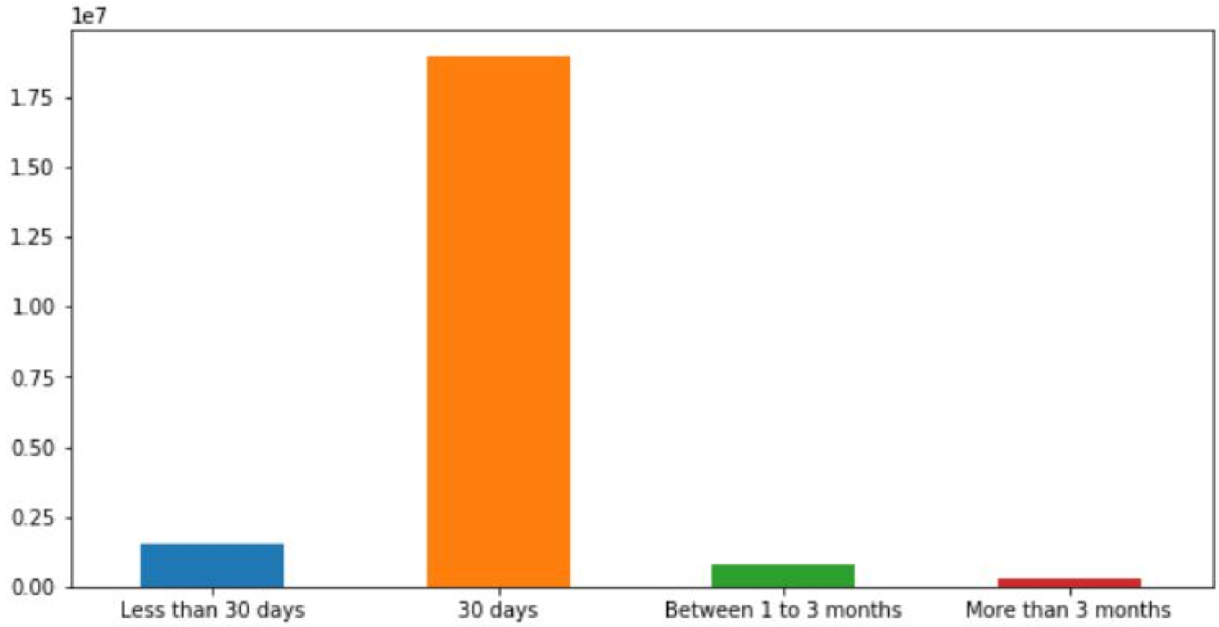


Figure 31: Distribution of the subscription lengths among KKBox users

Plan list price and Actual amount paid

Due to reasons unknown to the researchers, the subscription price may differ from the price that is actually paid by the user. As a matter of fact, out of 21,547,746 payment records, in 4% cases the subscription price was higher than the actual amount paid. Curiously enough, the vice versa scenario occurred in nearly the same amount of cases. Should these records state correct accounting figures and not be corrected in the future, the company lost 45,299,943 New Taiwan Dollar (1,468,193 United States Dollar on the 30/3/2019) as a result of these errors.

The following chart represents the distribution of subscription prices. It is clear that the most frequent subscription plans are at 149 New Taiwanese Dollar, 99NT$, 129NT$ followed by free subscription (presumably free trial period), subscriptions of the same or higher value than 150NT$.

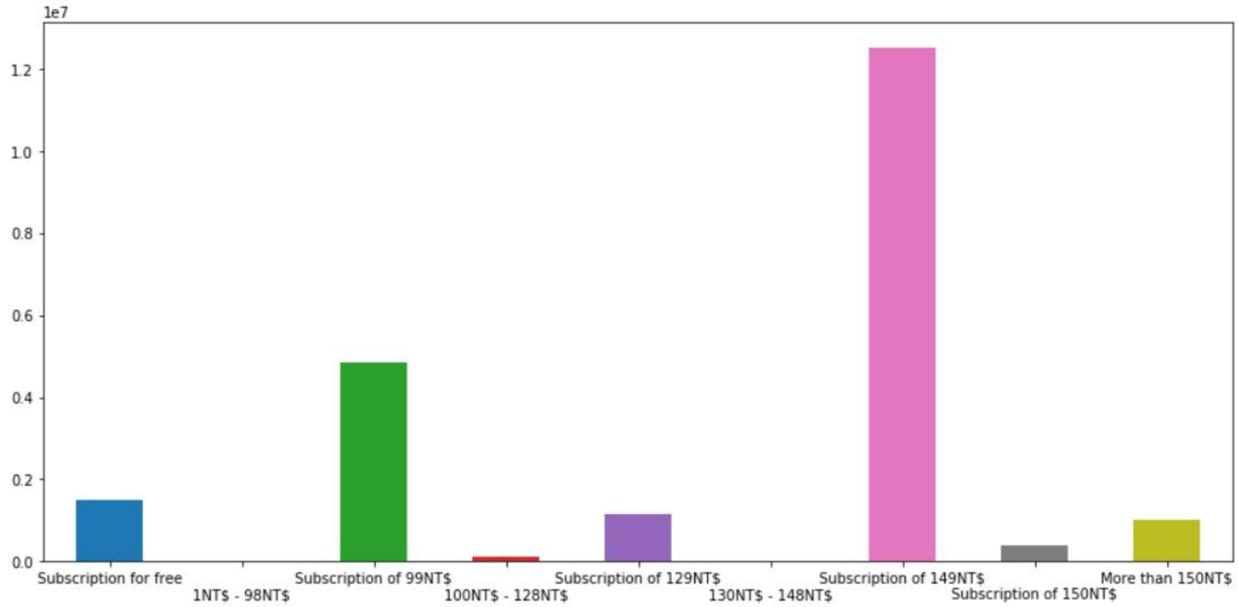


Figure 32: Distribution of the subscription prices among KKBox users

Deep Learning Input preparation

As the data in the Transactions dataset were recorded by the system, there are no faulty, or missing values. Nevertheless, the following steps were executed in order to prepare the dataset as an input for the Deep Learning architecture: outliers, aggregation together with feature engineering, dummy variables and feature scaling.

The first step included replacing outliers. In the case of Payment plan days, Plan list price and Actual amount paid, the outlying values were replaced by the highest value of non-outliers (mean value plus three times the standard deviation).

Transactions is a dynamic dataset, meaning a dataset with multiple values per user. Unlike in the User logs dataset where the researchers decided to exhibit a full period of two months including the missing days, in the case of Transactions, the researchers opted for aggregation. This decision was made due to the scarcity of timestamps within the selected time period. Usually there are between 1 and 2 timestamps within the period of 1st of January and the 28th of February, correspondingly with the usual subscription length of 30 days.

At the same time, the process of aggregation allowed extracting additional variables via feature engineering. For the ordinal variables of payment_plan_days, plan_list_price and actual_amount_paid supplementary variables were obtained, namely, count, first, last, min, max, sum and mean. This process for payment_plan_days, as conducted in the Python language, is shown in Figure 33. (note that the process is identical in the two other variables).

```
agg_plan_days = input_df.groupby(['msno']).agg({'payment_plan_days':["count","first","last",min,max,sum,"mean"]})
agg_plan_days.columns = agg_plan_days.columns.droplevel(0)

agg_plan_days = agg_plan_days.rename(columns = {"count":"subs_count",
                                "first":"subs_first",
                                "last":"subs_last",
                                "min":"subs_min",
                                "max":"subs_max",
                                "sum":"subs_sum",
                                "mean":"subs_mean",})
```

Figure 33: Code snippet: aggregating variable Payment Plan Days and employing feature engineering in order to derive further information

The description of each extracted feature is then available in the Table 5.

Extracted feature	**Feature description**
count	Number of timestamps per user
first	Value of user’s first timestamp
last	Value of user’s last timestamp
min	Value of user’s minimum timestamp
max	Value of user’s maximum timestamp
sum	Sum of all timestamps
mean	Average value of all user’s timestamps

Table 5: List of variables derived via feature engineering, including description

Dummy variables were created off the categorical variable of payment_method_id - which takes values between 1 and 41, meaning the dataset was extended by 41 columns - one for each value. The variables of is_auto_renew and is_cancel remained in their original form. Lastly, each column was scaled via min-max scaling in order to bring each value in the dataset to the range between 0 and 1.

The input shape of the Transactions dataset was 753856x64, as in 64 features about 753,856 unique users. In order to simplify, however, the Deep Learning architecture, the researchers decided to concatenate the Transactional dataset with Members dataset. Subsequently, the input shape of the merged dataset containing static user data was 753856x94. The concatenation process is shown in Figure 34:

```
multiinput_train = pd.concat([transactions_train, members_train], axis=1)
multiinput_test = pd.concat([transactions_test, members_test], axis=1)
```

Figure 34: Code snippet: concatenating the static datasets (Members and Transactions)

User logs

The User logs dataset is the bulkiest dataset with 392,106,543 rows providing insights into the daily user behavior of 5,234,111 users between the 1st of January 2015 and 28th of February 2017. Each day a customer uses the service, the following 9 features are recorded in form of a timestamp:

<table>
<tr><th>Variable</th><th>Variable type</th><th>Description</th></tr>
<tr><td>msno</td><td>Categorical</td><td>User id</td></tr>
<tr><td>date</td><td>Ordinal</td><td>Timestamp date</td></tr>
<tr><td>num_25</td><td>Continuous</td><td rowspan="5">Number of songs listened to
- 25%
- 50%
- 75%
- 98,5%
- 100%
of the song length on the particular day</td></tr>
<tr><td>num_50</td><td>Continuous</td></tr>
<tr><td>num_75</td><td>Continuous</td></tr>
<tr><td>num_985</td><td>Continuous</td></tr>
<tr><td>num_100</td><td>Continuous</td></tr>
<tr><td>num_unq</td><td>Continuous</td><td>Number of unique songs listened</td></tr>
<tr><td>total_secs</td><td>Continuous</td><td>Total period listened (in seconds)</td></tr>
</table>

Table 6: List of variables in the User logs dataset, including description

Features exploration and preprocessing

As the timestamps were recorded by the system, there are no missing values in the dataset. However, presumably due to possible errors in the software or a similar malfunction, the dataset contains outlying values which are impossible to achieve. List of the outlying values is as follows:

Feature	Minimum value	Maximum value
num_25	0	18798
num_50	0	1710
num_75	0	1690
num_985	0	2747
num_100	0	42004
num_unq	1	4784
total_secs	-9223372036854776.0	9223372036854776.0

Table 7: List of minimum and maximum values of the User logs variables

The outliers do not appear only in the total_secs column, where an error apparently occurred when recording the features, but also in the rest of the features. It is indeed unlikely that a single user would listen to thousands of songs or their respective proportions in a single day. The researchers therefore used Standard Deviation Method to detect the outliers and replaced them with mean value of the remaining values. This method creates the minimum and maximum threshold values within a feature through variation from the mean value. The cut-off value is usually three times the standard deviation of the feature. The following chart (Figure 35.) explains the method graphically.

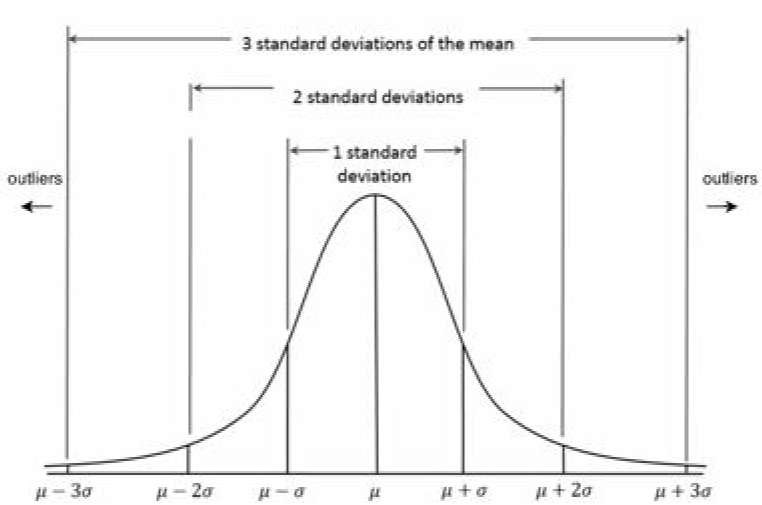


Figure 35: Visual representation of handling outlying values in a dataset via the Standard deviation method, Raj, 2016

However, due to the fact that the usual cut-off value of three standard deviations from the mean value covered too much of the sample (most likely due to the majority of entries being zero), the researchers extended the cut-off value to 50 times the standard deviation from the mean value. Any values lower or higher than this value were replaced by the mean value of the sample.

Deep Learning Input preparation

Once the dataset has been cleaned off the outlying values, the next step involved feature scaling. In this case, feature scaling serves not only the function to bring all the features to the same ground, but also to calculate the opacity of each pixel (one pixel = a single feature in a single day). As a result, each value has been brought on the scale between 0 and 1.

As mentioned before, the user logs features were provided in form of timestamps between the 1st of January 2015 and 28th of February 2017. In order to retain as much information as possible, the researchers decided not to aggregate the timestamps, but rather present them as a whole across a selected period. It is noteworthy, as explained later in the chapter *Train and Test split*, that the researchers opted not to use the whole timeline starting in the beginning of 2015, but selected only the first two months of 2017. A longer period of time might distort the Machine Learning capabilities and would naturally exclude or skew the most recently registered users. However, as only the days with any sort of activity from the user are recorded, there are missing days from the whole timeline in the dataset provided by KKBox and each user thus has a different shape of the matrix representing their behavior. The usual padding methods of both pre-sequence padding and post-sequence padding would distort the chronological order as the matrix would be firstly filled with non-empty timestamps and later with padding from top to bottom. The objective was therefore to chronologically fill out missing days with zeros in between actual timestamps. This way, each user's activity was represented with a matrix of 59 days chronologically representing their activity within the period of 59 days between the 1st of January and the 28th of February.

The final step in preparing the input for the Deep Learning architecture was to convert these matrices into heatmaps for Image recognition, the reasons for choosing Image recognition at the expense of Time series analysis are explained in the chapter *Churn analysis using image recognition*. The following code snippet examines the process of converting a file of matrices into heatmaps.

```
def processInput(i):
    with h5py.File('user_logs_test.h5', 'r') as hf:

        user = i
        data = hf["user_logs"][user*days_no:user*days_no+days_no]
        data = data.reshape(7*days_no)

        pixel_list = [(int(round(data[i]*255))) for i in range(len(data))]
        pixel_list = np.array(pixel_list)
        pixel_list = pixel_list.reshape(days_no,7)
        pixel_list = [(int(round(data[i]*255)),int(round(data[i]*255)),int(round(data[i]*255))) for i in range(len(data))]

        ratio = 10 #rescaling the image

        name = str("{0:0=6d}".format(i))

        img = Image.new('RGB', (7, days_no))
        img.putdata(pixel_list)
        img = img.resize((7*ratio,days_no*ratio))
        img.save('DeepLearning_input/userlogs_test/'+name+'.png')
```

Figure 36: Code snippet: transforming matrices of user behavior into heatmaps

The following two figures show the final representation of the input for the Deep Learning architecture. Each heatmap has the dimensions of 7x59 representing 7 above-mentioned features (excluding msno and date) across 59 days. The activity heatmap on the left (Figure 37.) belongs to a non-churning customer where a significant activity is apparent in the right part of the image. On the other hand, a churning customer is shown in the Figure 38. on the right with no visible activity.

Figure 37: Heatmap representing user behavior: activity of non-churning customer

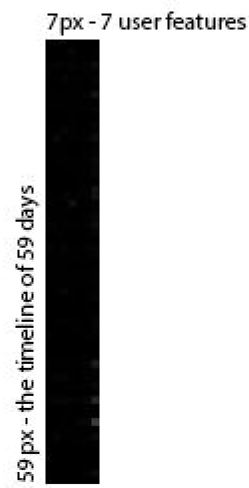


Figure 38: Heatmap representing user behavior: activity of churning customer

Customer churn prediction model

The following chapter describes in detail the application of knowledge acquired via literature review in chapter *Deep Learning Customer Churn model*. Firstly, the researchers need to adjust the inputs in order to equalize the class distribution, meaning that for training purposes, the proportion of churners and non-churners in the input dataset ought to be balanced. Next, the train and test split described in chapter *Train and Test split* is conducted as a prerequisite step for the Deep Learning training. Subsequently, the construction of the Deep Learning architecture is thoroughly described, together with the training process and evaluation metrics. In order to derive as much marketing insight as possible given the data available, the researchers employ interpretation model which yields churn predictors used for each customer's churn probability. As a result, the entire chapter serves as an exploration how could customer data from a subscription-based company be employed in a Deep Learning Customer Churn Prediction model and what marketing insights could be derived from such a process.

Imbalanced class distribution

In the case of the KKBox dataset, narrowed down to the aforementioned period between 1st of January - 28th of February 2017, the class distribution heavily favors non-churning customer over the churning ones with 93% and 7% of the dataset, respectively.

Despite the fact that the researchers would have far smaller sample, they were forced to use under-sampling method. As far as the information gathered reaches, over-sampling of the heatmaps representing customer activity was not reliable enough to opt for this method. The researchers therefore selected all the churning customers (49,553) and randomly picked 50,447 of non-churning customers to equalize the class distribution in the dataset. As a result, the entire dataset for training and testing comprised 100,000 users.

Train and Test split

The researchers have at their disposal three datasets, each of them consisting of different type of information about 100,000 users. The researchers decided to split the datasets into a training and testing datasets typically for Machine Learning in 80/20 ratio with 80% of each dataset for training and remaining 20% for testing. The training subset contains known label of each customer (non-churn/churn) while the testing dataset serves to test the capabilities of the trained model on data with hidden label (Bronshtein, 2017).

Deep Learning Architecture

Description

Based on the aim of the classification model and data available, the researchers drew inspiration from two related works mentioned in the chapter *Related work to the KKBox Deep Learning architecture*. Two important aspects were crucial when building the Deep Learning architecture. First, there are multiple data streams which would be deprived of their explanatory power if concatenated together. Second, the latest research explains that transforming customer behavior into heatmaps and using image recognition classifier yield far more satisfactory results than previous methods. That said, the researchers built a two-input Deep Learning architecture via Keras Functional API.

Unlike Keras Sequential models, Keras Functional API allows researchers create models with multiple inputs as well as more flexibly define individual layers. With this approach, model input is defined with specified shape of the input dataset in the first step. Second, the functional core of the model is created by connecting layers with designated functions to each other. As the last step, the model is created by defining the input and output layers (Brownlee, 2017). The building process of the Deep Learning architecture is described in Figure 39:

```
# Members + Transactions
static_input_3 = Input(shape=(94,), name='static_input_3') # input layer
dnn_out_1 = Dense(units=94)(static_input_3) # DENSE_1
dnn_out_1 = Dropout(rate = 0.25)(dnn_out_1)
dnn_out_1 = BatchNormalization()(dnn_out_1)
dnn_out_1 = Activation('sigmoid')(dnn_out_1)

dnn_out_2 = Dense(units=94)(dnn_out_1) # DENSE_2
dnn_out_2 = Dropout(rate = 0.25)(dnn_out_2)
dnn_out_2 = BatchNormalization()(dnn_out_2)
dnn_out_2 = Activation('sigmoid')(dnn_out_2)

dnn_out_3 = Dense(units=94)(dnn_out_2) # DENSE_3
dnn_out_3 = Dropout(rate = 0.25)(dnn_out_3)
dnn_out_3 = BatchNormalization()(dnn_out_3)
dnn_out_3 = Activation('sigmoid')(dnn_out_3)

dnn_out_4 = Dense(units=94)(dnn_out_3) # DENSE_4
dnn_out_4 = Dropout(rate = 0.25)(dnn_out_4)
dnn_out_4 = BatchNormalization()(dnn_out_4)
dnn_out_4 = Activation('sigmoid')(dnn_out_4)

dnn_out_5 = Dense(units=94)(dnn_out_4) # DENSE_5
dnn_out_5 = Dropout(rate = 0.25)(dnn_out_5)
dnn_out_5 = BatchNormalization()(dnn_out_5)
dnn_out_5 = Activation('sigmoid')(dnn_out_5)

# User logs
userlogs_input_3 = Input(shape=(59,7,3), name='userlogs_input_3')
input_ = Dropout(0.2)(userlogs_input_3)
input_ = Conv2D(64, kernel_size = (7,1,), padding='same', activation='relu', strides=1)(input_)
input_ = Conv2D(64, kernel_size = (1,7,), padding='same', activation='relu', strides=1)(input_)
input_ = Flatten()(input_)

# User_logs + Transactions + Members
merged = keras.layers.concatenate([dnn_out_5,input_]) # Merged layer
combined_out = Dense(2)(merged)
combined_out = Activation('softmax')(combined_out)
```

Figure 39: Code snippet: creating a multi-input Deep Learning architecture in Python

The architecture used in the KKBox Customer Churn analysis therefore comprises two input branches - one for static data, namely concatenated datasets of Members and Transactions, and one for dynamic data, heatmaps representing customer behavior. The former branch, specifically Fully connected Neural Network, consists of the input layer of the shape (94,) meaning any number of rows times 94 features. This shape of input must be identical to the number of features in the dataset that is parsed into the model. The input layer is followed by consequent five sets of Dense, Dropout, Batch Normalization and Activation layers. While Dense layer is a regular Neural Network fully connected layer, Dropout layer randomly turns off certain units which regulates the networks and prevents overfitting (Keras.io, n.d.; Ketkar, 2017). The Batch

Normalization layer speeds up the training process and allows other layers to be more independent of one another thanks to normalizing the values in hidden layers (Doukkali, 2017). Lastly, Activation layer decides which neurons to be activated (Sharma V, 2017).

The second branch which takes an input of the user activity represented as heatmaps is a Convolutional Neural Network. The input shape is (59,7,3,), meaning 59 pixels in height representing the time period of 59 days, 7 pixels in width as 7 user features described previously, each pixel having 3 channels for R, G and B color proportions. The input is then connected to a Dropout layer and two Convolutional layers. The first Convolutional layer has kernel size of 7x1 and its purpose is to find weekly patterns in a single feature. The second Convolutional layer consists of kernel size of 1x7 and it attempts to find traits in one day across all the activity features. The following Flatten layer converts the input into a desired output dimensionality (Ketkar, 2017).

Lastly, both of the branches are merged into a binary classification output thanks to Concatenate layer. The structure of the architecture used for KKBox customer churn classification is plotted in Figure 40. Due to the length of the original model summary, the figure has been split horizontally for readability reasons.

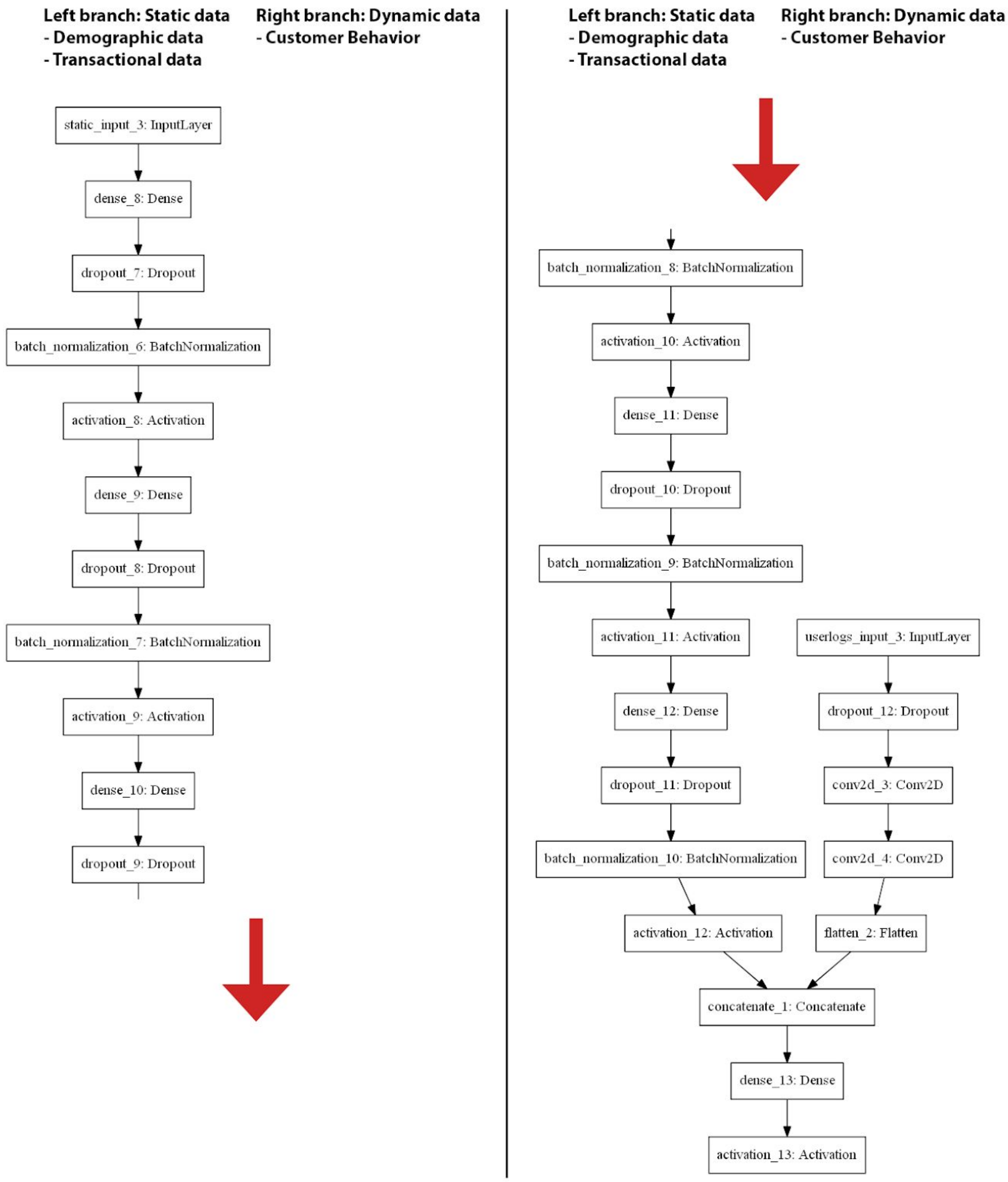


Figure 40: Deep Learning model summary (split for readability reasons)

Training process

As mentioned in the chapter *Train and Test split*, the model was trained on 80,000 samples with the class distribution of 49,55% of churning customers and 50,45% non-churning customers.

The training of the model was conducted over 50 epochs (repetitions) with an epoch being passing forward and backward of the entire dataset once. The number of epochs is not a one-fits-all parameter and it differs from case to case. Thus, the number of epochs was determined by the point when the model's learning curve began to stagnate. The entire dataset was parsed through each epoch in batches of size 500 which results in 160 iterations (size of the training dataset 80,000 divided by the batch size of 500) (Sharma, 2017).

During training, the model's accuracy was validated on the testing dataset. This provides two sets of metrics during the training - accuracy and loss for both training and validation dataset (Brownlee, 2016). While accuracy is the ratio of the number of correct predictions and number of total predictions, loss penalizes the wrong classification and is usually used for multi-class predictions (Mishra, 2018).

The learning curve of the model's accuracy is described in Figure 41. The blue line shows the progression in learning with validating on the training date while the orange line shows the learning curve validated on the testing data. The comparison of the two lines indicates that even though it is easier to predict on the training dataset, the model still learns over epochs for both datasets. Model's accuracy by the 50th epoch is 82,84% on the training dataset and 74,91% on the validation (testing) dataset.

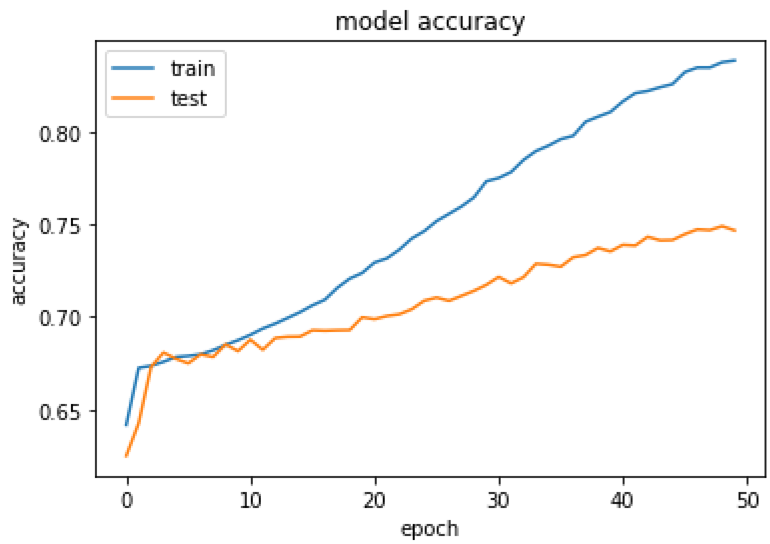


Figure 41: Model training - learning curve: model accuracy

On the other hand, the following Figure 42. represents the curve for the loss function. A loss function is the plotting of the result's deviation from the actual classes and is used to optimize the model (Parmar, 2018). Once again, the blue line represents the loss function for the training dataset while the orange one indicates the same metric for the testing dataset.

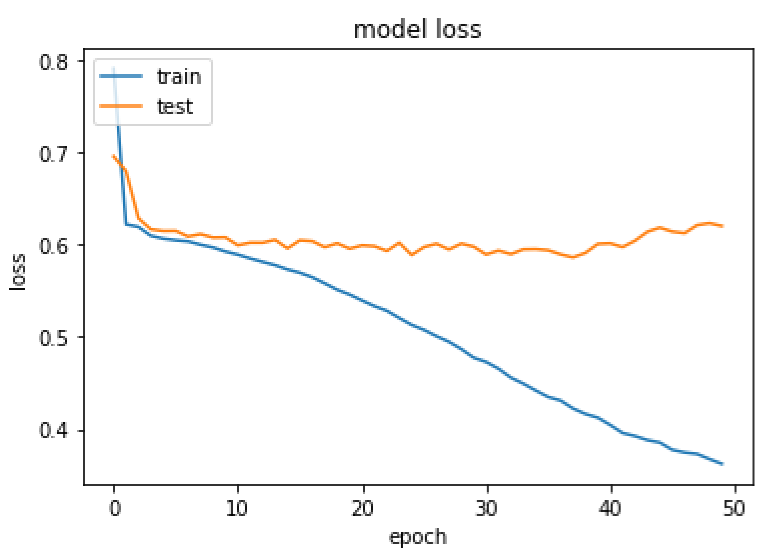


Figure 42: Model training - learning curve: model loss function

The comparison of the two curves unfortunately expresses signs of overfitting after the 10th epoch. The issue of overfitting indicates that the model learned too well on the training dataset and included statistical noise or randomness for training. The problem can occur if the model is trained for too long or if it is more capable than required for the problem. Such a model has difficulties generalizing to new datasets (Brownlee, 2019).

Evaluation metrics

Despite the loss curve suggesting an overfitting problem, the ultimate method how to assess a model's predicting capabilities are evaluation metrics presented in this chapter.

Confusion matrix is a crosstab of both predicted and true classes of the testing dataset. The terms used to describe the overlaps between the predicted and true values are described in Table 8:

	Predicted 0	Predicted 1
Actual 0	**True negatives** (non-churning customers classified correctly)	**False positives** (non-churning customers classified as churning)
Actual 1	**False negatives** (churning customers classified as non-churning)	**True positives** (churning customers classified correctly)

Table 8: Confusion matrix representing binary-classification results

As stated before in chapter *Train and Test split*, the testing dataset contains 20,000 users. The confusion matrix is therefore as follows:

```
Predicted      0     1    All
Actual
0           7629  2445  10074
1           2573  7353   9926
All        10202  9798  20000
```

Figure 43: Confusion matrix of the KKBox Deep Learning architecture

On the horizontal axes (rows), the crosstab shows Actual labels with 0 standing for non-churners and 1 for churners. The vertical axes (columns) represented classifications made by the model. The testing dataset therefore contained 10,074 non-churning customers out of whom the classification model predicted 7,629 correctly. On the other hand, out of 9,926 actual churners, the model predicted 7,353 correctly. Furthermore, these values serve as inputs for the calculation of the afore-mentioned accuracy, as shown in Figure 44. (Mishra, 2018).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Figure 44: Equation for calculating Deep Learning model's accuracy, Google Developers, 2019

In this case the accuracy is calculated as (7,353+7,629)/20,000 = 74,91% which equals the model's accuracy by the 50th epoch as shown in the chapter *Training process*. These metrics, as validatory as they seem, are, though, not enough to truthfully assess the performance of a model. Therefore, further metrics ought to be deployed in order to evaluate other aspects of the classification model.

The values in the confusion matrix serve as a foundation for additional metrics. True Positive Rate (in other words Sensitivity or Recall) which suggests the proportion of correctly classified churners (calculated as shown in Figure 45.) is 74,08% in the case of KKBox churn classifier, meaning nearly 3 out of 4 churning customers were classified correctly.

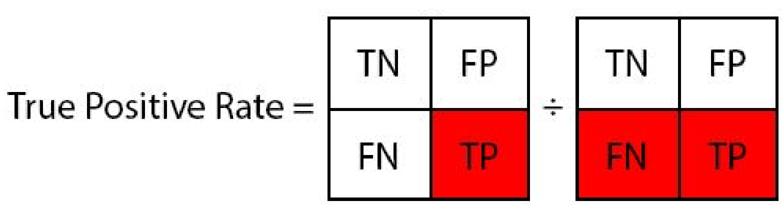


Figure 45: Visualization of True Positive Rate equation

On the other hand, False Negative Rate (or Specificity) is conversely calculated as shown in Figure 46. and expresses the proportion of correctly classified non-churners. The models results presented in the confusion matrix above suggest 75,73% Specificity ratio. Consequently, False Positive Rate is defined as 1 - Specificity or as FP/(TN+FP). In either case, FPR equals 24,27%.

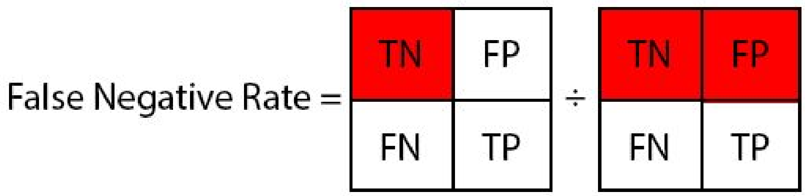


Figure 46: Visualization of False Negative Rate equation

The two above-mentioned metrics (True Positive Rate and False Positive Rate) are the bases for other metrics used to evaluate a model's performance, namely Receiver Operator Characteristics and Area Under the Curve. The ROC curve plots the TPR and FPR at various threshold settings between 0 and 1 other than just the usual split of 0.5 where the probabilities higher than 0.5 are considered a Positive (1 - churner) and the probabilities lower or equal 0.5 are considered a Negative (0 - non-churner). A perfect model plotting 100% of True Positives and 0% of the False Positives is plotted in Figure 47. (Ryckel, 2017; Narkhede, 2018). It is noteworthy that this chart represents the best-case scenario which is impossible in the real-market setting.

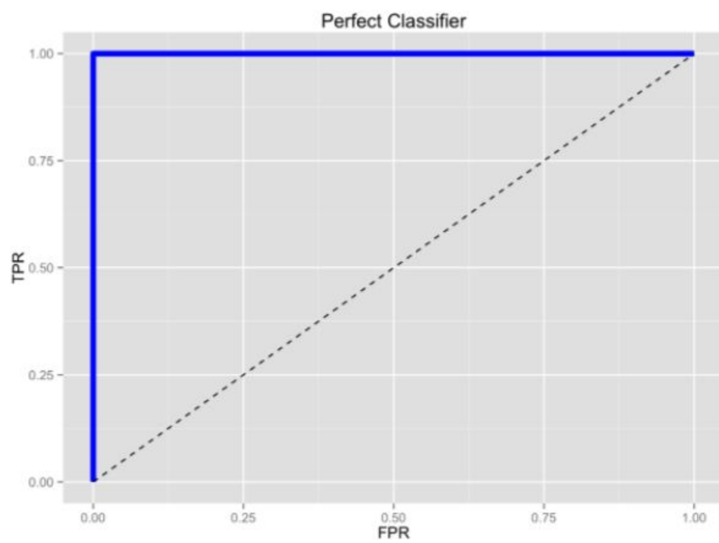
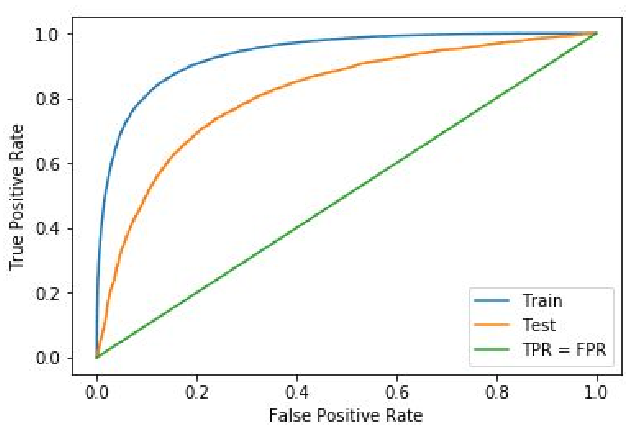


Figure 47: The plot of Receiver Operator Characteristics in the best-case scenario, Ryckel, 2017

The consequent Area Under the Curve then suggest how good the model is at differentiating between the positive and negative class. The higher the ratio on a scale from 0 to 1, the better performing model. The Receiver Operator Characteristics and Area Under the Curve are present in Figure 48.



Area Under the Curve (AUC): 0.8128186099470807

Figure 48: Receiver Operator Characteristics including Area Under the Curve with regard to KKBox Deep Learning model's results

The last two metrics worth calculating are Precision and F1 score. Precision explains how many of the total number of predictions made for each class were actually correct. The formula is TP/Total predicted Positive or TP/TP+FP (Ping Shung, 2018). The Precision for non-churner predictions is 74,78% (7629/(7629+2573)) and for churner predictions 75,05% (7353/(7353+2445)). The F1 score finally provides an insight about the balance between Precision and Recall. This is an important metric for imbalanced business datasets where a multitude of True Negatives (non-churner; see chapter *Imbalanced class distribution*) can contribute to high accuracy of the model, yet lead to business costs related to False Positive and False Negative predictions (Ping Shung, 2018).The formula for the F1 score is described in Figure 49.

$$F1 = 2 * \frac{1}{\frac{1}{precision} + \frac{1}{recall}}$$

Figure 49: The equation for F1 score, Mishra, 2018

Finally, the F1 score for either class is 75%. The classification report summarizing the above-mentioned metrics is shown in the Figure 50.

	precision	recall	f1-score	support
0	0.75	0.76	0.75	10074
1	0.75	0.74	0.75	9926

Figure 50: Metrics summary

In marketing terms, the confusion matrix (Figure 43.) together with Precision would be ultimately the most interesting metric. If this model would be the only input as to who to target in the retention strategy and no other criteria would come into picture, on this particular sample of 20,000 users, KKBox would intervene 9,798 users. Out of them, 7,353 would be targeted correctly, meaning 75,05% (Precision) were actually churning customers who were targeted as part of the retention strategy. The remaining 2,445 customers were targeted as churners despite not having any intentions of churning. These users represent cost in a way that offer is tailor-made for them and frequently this offer is discount based. Therefore, these customers were offered a price deduction even though they were not going to leave KKBox. On the other hand, False negatives represent 2,573 customers who were going to churn but were not detected by the prediction model. This creates an opportunity cost where interventions were not employed in cases where they should have.

Predictions Interpretability

In order to interpret the prediction decisions and derive further insights for forming eventual marketing retention strategies, the researchers opted for Local Interpretable Model-agnostic Explanations or LIME, described in chapter *Interpretability of Deep Neural Networks*. The aim of LIME's utilization was to see what features in particular the customer churn prediction model used to assess the probability of non-churn / churn of each user. These features would lately serve as an additional insight for understanding reasons behind a customer's churn.

However, after a thorough research and discussion with the author of the interpretation model for Python programming language, it has been concluded that a Deep Learning architecture as complex as the one used for KKBox churn classification, cannot be yet interpreted by the model. The author Marco Tulio Ribeiro from the University of Washington, responded on the researchers' request saying that "There is no way to have a 3d tabular input, as I don't know how to perturb that," (GitHub, 2019).

As a result, the researchers were forced to choose a workaround in form of splitting the Deep Learning architecture in each respective branch solely for the purposes of interpretation. One branch is used for the static data (concatenation of Members and Transactions) and the other half is used solely for image classification of the heatmaps representing users' activity. Each branch had to be trained separately and consequently used merely for predictions.

The outcome of the interpretation model for the static data branch is shown in Figure 51. In this particular case, the prediction model classified user as churn with 90% probability. The features used for this classification were Registration method 7, Registration method 4 and City 1 for churning and Registration method 3 and Registration method 9 for non-churning.

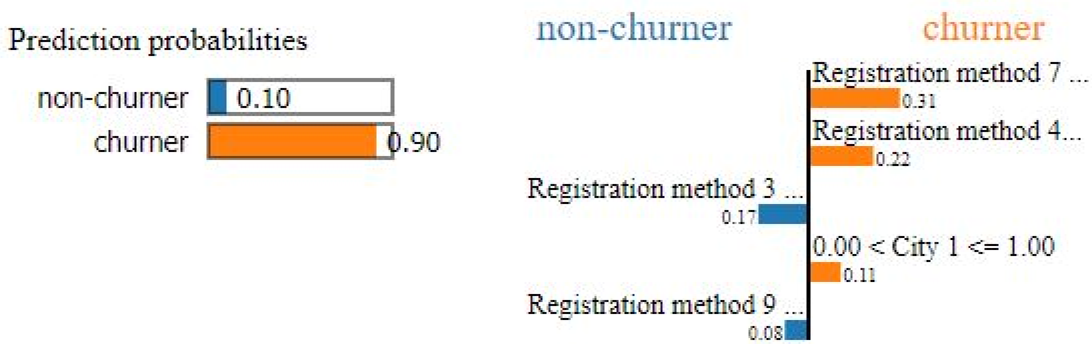


Figure 51: Predictions interpretability of the static branch

On the other hand, the interpretation model for the dynamic dataset yield results that are present in Figure 52. While the chart on the left is the result of the LIME model, the chart on the right shows actual user activity. As per the highlighted area, the model made the decision based on the activity around two-thirds down the time period, therefore around the first week in February.

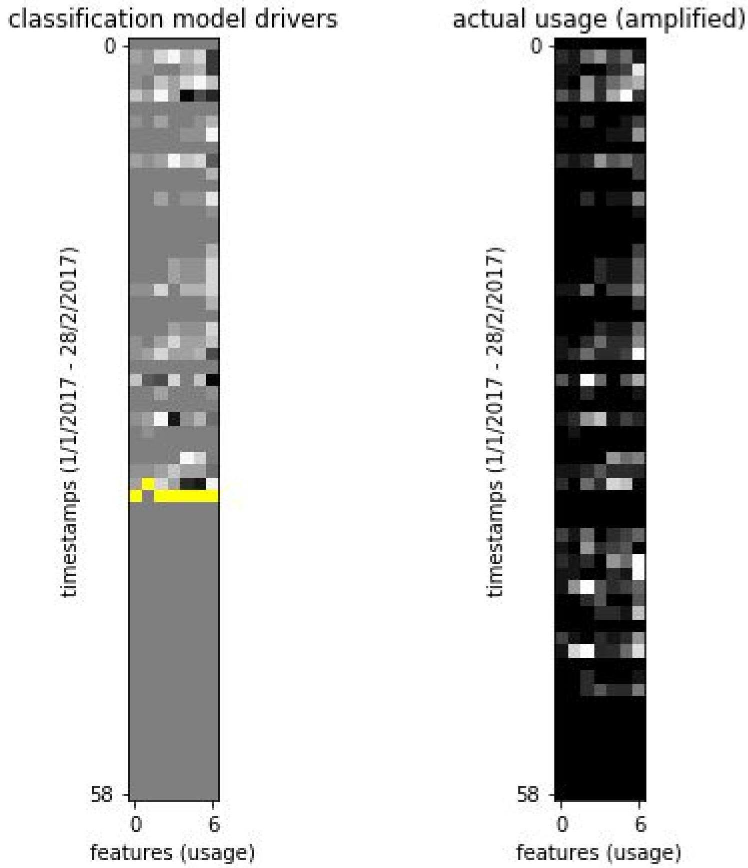


Figure 52: Predictions interpretability of the dynamic branch (user activity heatmaps)

Despite this approach being the only possible workaround known to the researchers, there are clearly drawbacks which might jeopardize understanding reasons for customer churn if relied upon too heavily.

Firstly, as the branches were trained separately, there is naturally difference between the predicted classes not only by the two individual branches, but also between them and the two-input model. This issue spans across the interpretation as well; it is reasonable to believe that the two-input prediction model used different features and reasoning for the predictions than the ones presented by the individual interpretation branches. Secondly, as seen from the examples, the results yielded from the model do not provide marketers with much clearer image as what the actual churn drivers might be. This issue is rooted in the discrepancy between the available input data for the KKBox case and possible churn drivers, as seen in the literature chapter *Drivers of customer churn*. It is noteworthy to say that this particular downside is data-bound, therefore the problem may diminish with other source data.

As a result, despite the classification model predicting with high accuracy, the interpretations provided by LIME in this particular case should server merely as an estimation and should be retroactively checked with the initial data points.

Synthesis of the Conceptual Framework and Customer churn prediction model

The following chapter presents the applicable summarization of the above-described results in the context of the framework for CRM Retention Strategies. It is, though, noteworthy that the initial purpose of the data used by the researchers to construct the Deep Learning model was not to cover the entire topic of creating CRM retention strategies, but to merely predict customer churn. As a result, the capabilities tied to the data provided are fairly limited and unfortunately cannot cover the Conceptual Framework in its entirety.

The following Figure 53. represents the Conceptual Framework from the chapter *Theoretical conceptualization of designing CRM retention strategies,* with the impact of results of the Deep Learning Customer Churn Prediction model. The framework highlights in red the parts which can the results cover completely and in yellow the ones which are covered just partially. These parts will also be addressed on the following pages. It is also worth noting that due to the highly interconnected nature of the respective parts of the framework, the model may have indirect influence on other parts as well, but these will be explained only briefly.

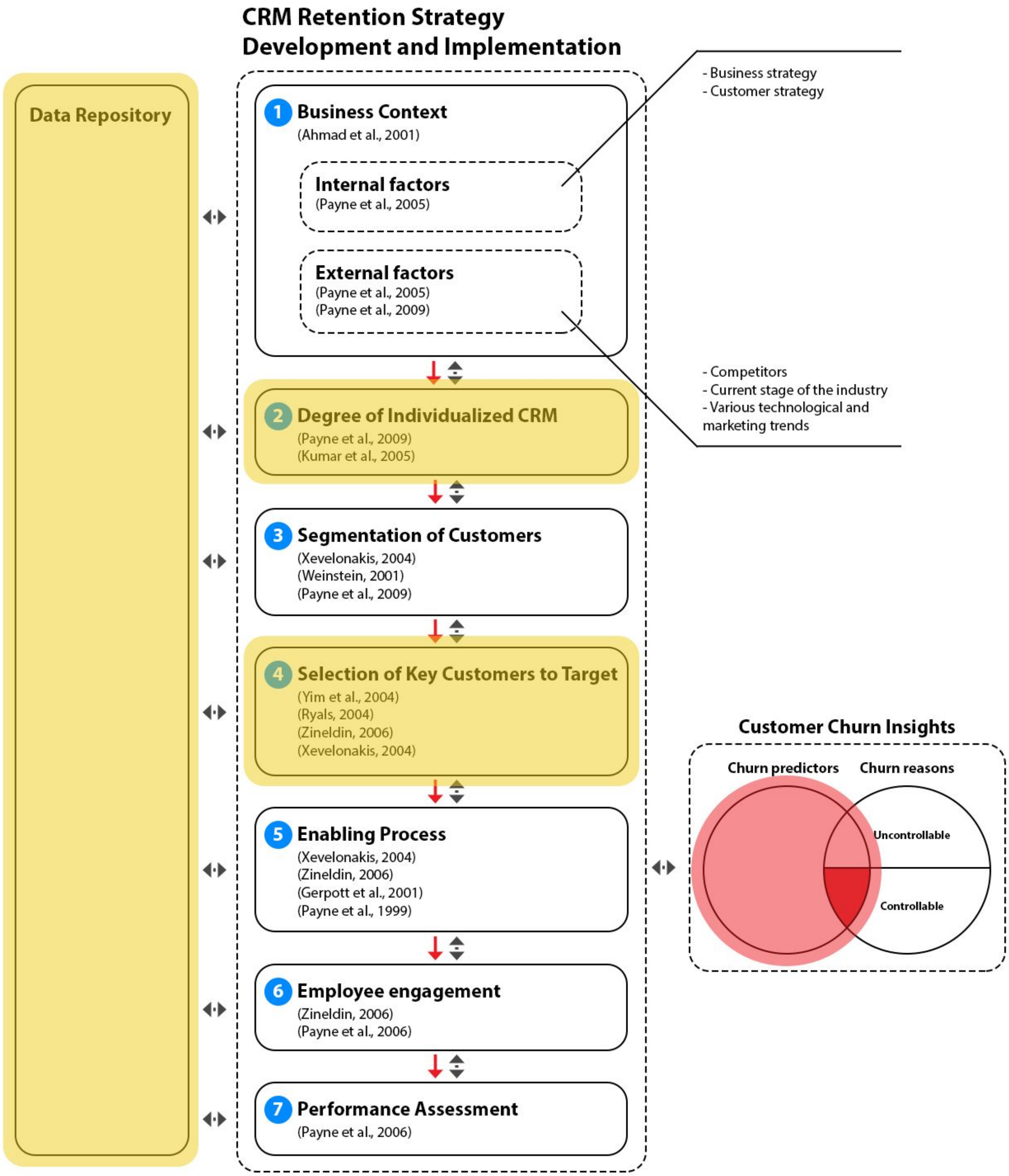


Figure 53: The influence of Deep Learning Model Predicting Customer Churn on the Self-Developed Conceptual Framework of the Process of Development and Implementation of CRM Retention Strategy

Lastly, the non-highlighted parts represent the areas where:

- The researchers lack additional data needed for further analysis
- The Deep Learning Customer Churn Prediction model has no impact (with the exception of Performance Assessment)

In the best-case scenario, the information derived from the Deep Learning model would be supplemented by the following insights to create the most complete foundation for the CRM Retention strategy. First and foremost, these features include having a clear vision of the business and customer strategy which make up the internal factors as well as having examined the external factors such as the stage of the industry, information about competitors or other macro-economic factors. Other prerequisite data in order to fully cover the conceptual framework include the branding standards, multi-channel means and choice of tactics and actions as part of the Enabling process.

Consequently, the next step would be to fully inform and train employees who are in touch with the customers so that their eventual retention endeavours are in line with the goal of the strategy. Lastly, keeping a record of all the customer interventions within the CRM Retention Strategy is an imperative task in order to assess the success of the campaign and to use the information for future reference. The information derived from this part of the framework could potentially lead to another analysis adding to “Selection of Key customers” where based on the historic information, the response probability to customer intervention is calculated. All of these features provide additional bases, on top of the Deep Learning Customer churn findings, for a more informed and complete CRM Retention Strategy.

On the other hand, the following parts of the Conceptual Framework can be either fully or partially enhanced by the Deep Learning Customer Churn Prediction model. The *data repository* is a first step that is enriched by the Deep Learning model. As the model produces customer churn prediction and predictors of churn on the individual level, companies acquire new information about their customers which can be used in the processes of CRM retention strategy.

This closely influences the *degree of individualized CRM* where if company attempts to shift towards individualized CRM, the more complete picture of information about each one of their customers enables this choice. However, there is also high possibility that the decision of creating a DL CCP model is a result of choosing individualized CRM, where more information on individual-level of customer is needed. Due to the limited information about KKbox, the researchers are not able to decide what degree of individualized CRM should be made, however it is fair to assume that with the decision of predicting customer churn on individual level, KKbox leans towards the individualized CRM.

The next step that is being influenced by the predictions of the customer churn on the individual level is the *selection of key customers to target*. This is also arguably the main reason why the Deep Learning model is made. As the literature shows, there are two main approaches to this selection. Firstly, the approach of maximizing profits presented by Lemmens et al. (2017) which bases this selection on the probability of customer churn, expected residual CLV, response probability and incentive cost. The second approach, aiming to minimize churn, is presented by Ascarza (2018), which bases the selection only on the response probability. Since the customer churn probabilities are calculated, the approach of Lemmens is more appropriate in order to use the results of the Deep Learning model.

Furthermore, the predictors of churn is the only step in the conceptualized framework that is being fully covered by the data from the model. However, in order to use this information for marketing purposes, companies need to distinguish which of these predictors are also controllable drivers of churn. The marketers responsible for building the CRM Retention Strategy need to find an overlap between the churn predictors yielded from the Deep Learning model and the actual controllable churn reasons. If they manage to do so, they can take advantage of these insights for a tailor-made offer based off the customer’s reasons to churn (eg., if an individual’s reason was an unaffordable price, the intervention ought to be built around a discount offer).

In order to derive as much marketing insight from the data available as possible, the researchers opted to calculate the CLV of the KKBox customer base from the information in the Transactions dataset. The CLV feature is one of the prerequisite pieces of information when conducting selection of key customers. It is, however, important to point out that the data provided during the KKBox competition are not primarily aimed at assessing customer value for retention campaigns. Nevertheless, the researchers attempt to calculate customer worthiness while understanding that the following approach is heavily limited by the data provided and does not meet the criteria set in the conceptual framework present in *Theoretical conceptualization of designing CRM retention strategies*.

The essential calculation of Customer Lifetime Value is the difference between profit generated by a customer and costs linked to acquiring and maintaining that customer. The calculation used in this analysis is based on the following premises:

- The data available is extremely limited providing only the monetary value of each subscription.
- The researchers assume that even though the data lacks information regarding KKBox costs, the individual cost linked to a customer would not differ or be easily distinguished from another customer's costs.

The Customer Lifetime Value for each customer is therefore calculated as the sum of all of the subscriptions' monetary value paid. In order to put each customer's value into perspective, a percentile is provided. The outcome of the calculation is shown in Figure 54.

```
Customer value: 540NT$ which is more or equal than 16.05% of KKBox customer base.
```

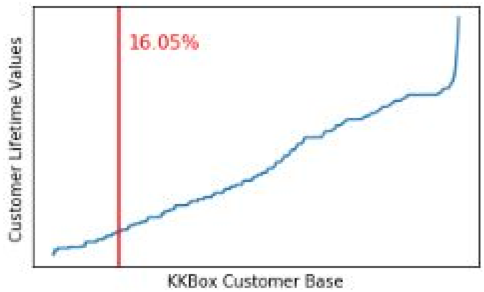


Figure 54: An individual's Customer Lifetime Value in the context of the entire KKBox customer base

The scatter plot represents all the 20,000 users within the sample on x-axis and their respective values on y-axis with the minimum Customer Lifetime Value being 0NT$ (New Taiwanese Dollar) and the maximum 8,138NT$, meaning that the customer with the highest CLV paid KKBox throughout his or her entire tenure 8,138NT$ (approximately 230€). The customer in Figure 54. has CLV of 540NT$ which is more than 16,05% of the 20,000 KKBox customers.

Marketing application

The Deep Learning classification model provides marketers or data analysts with the list of all the predicted values out of which the marketer can select only the churning ones and individually assess their probability of churn together with possible churn drivers and Customer Lifetime Value in the context of the entire customer base.

The entire dashboard for a single customer who has been classified as churner is shown in Figure 55. The panel provides information on an individual level which can be utilized for retention strategy. User ID which could be linked to other Customer Relationship Management tools, is followed by individual Customer Lifetime Value. The value is calculated and presented in the context of the entire customer base, as described in the previous chapter. The user below paid 540 New Taiwan dollars during the entirety of his tenure with KKBox which puts him on 16th percentile, meaning he has paid more or equal than 16,05% of the whole KKBox customer base.

Next, the prediction probabilities are present and consequently translated into a class based on a 50% threshold. In this particular case, the prediction favored churn class with more than 99,97% probability. The actual class in this case serves merely as confirmation of the accuracy of the model. In real-market predictions, no actual class is available. Lastly, predictors of customer churn and their weights are derived for both static and dynamic input. The discrepancy in the predictions probability between the static model and the full model is due to the difference in input as explained in chapter *Predictions Interpretability*. These predictors may provide valuable insight for understanding the reasons behind churn (or non-churn) of an individual customer. However, as mentioned before, these are to be used with caution.

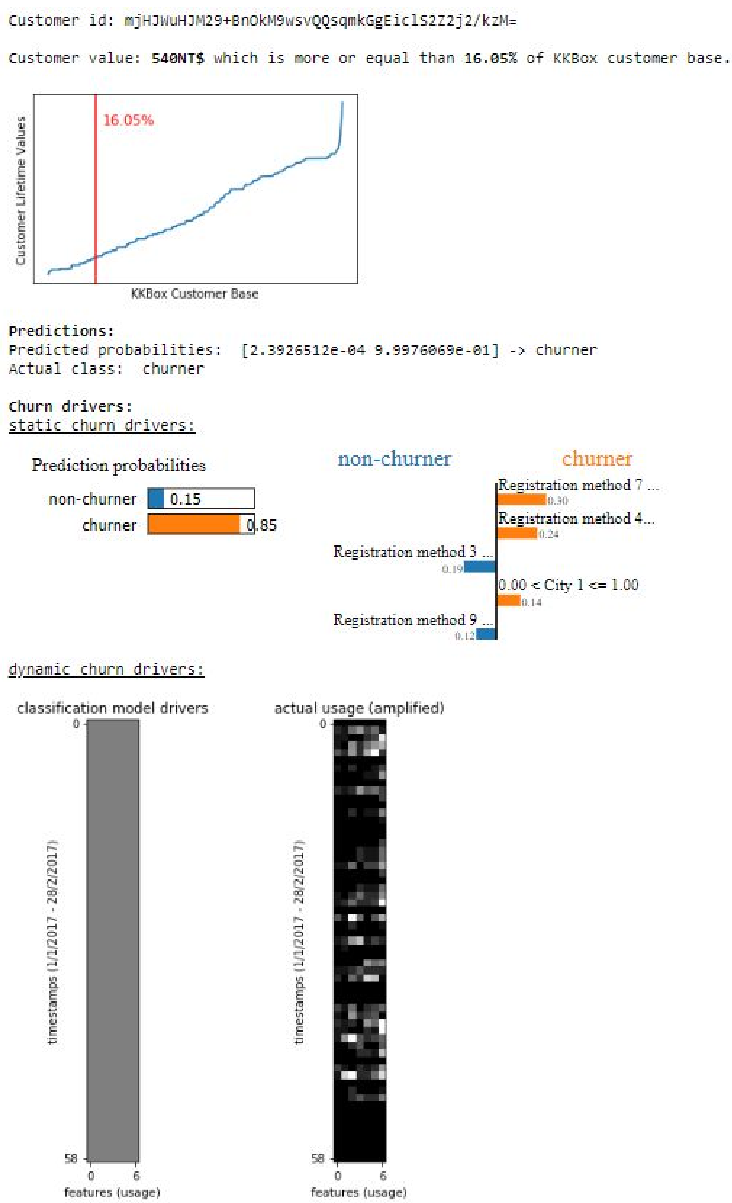


Figure 55: Managerial dashboard for an individual KKBox customer (churning customer)

Part 1: Customer Lifetime value
The CLV of the individual customer in this dashboard is 540NT$ (the amount of money he or she paid throughout the whole tenure with KKBox). That means that the customer has paid more than 16.05% of the KKBox customer base.

Part 2: Churn probability
The prediction model assessed this customer as churner with 99,97% probability (the customer was classified correctly)

Part 3: Churn predictors
The prediction model classified the customer as churner based on the following features in the static branch:

- Registration method 7
- Registration method 4
- City 1

At the same time Registration method 3 and 9 were favoring non-churning class.

As for the dynamic features, meaning customer behavior, the model does not provide any visible clue, presumably due to the lack of activity.

On the contrary to the previous example, the dashboard present in Figure 56. represent a customer who has been classified as non-churner. User is at 78th percentile which means he is among the 22% most profitable KKBox's customers. In this particular case, the probability of churn according to the Deep Learning model is around 2%. Even though four out of five features most contributing to the prediction stand by churn, Age and the culmination of other predictors below top five over-weighted the prediction in favor of the non-churn class. At the same time, user seems to have been actively using the service, which is another aspect contributing to the final prediction.

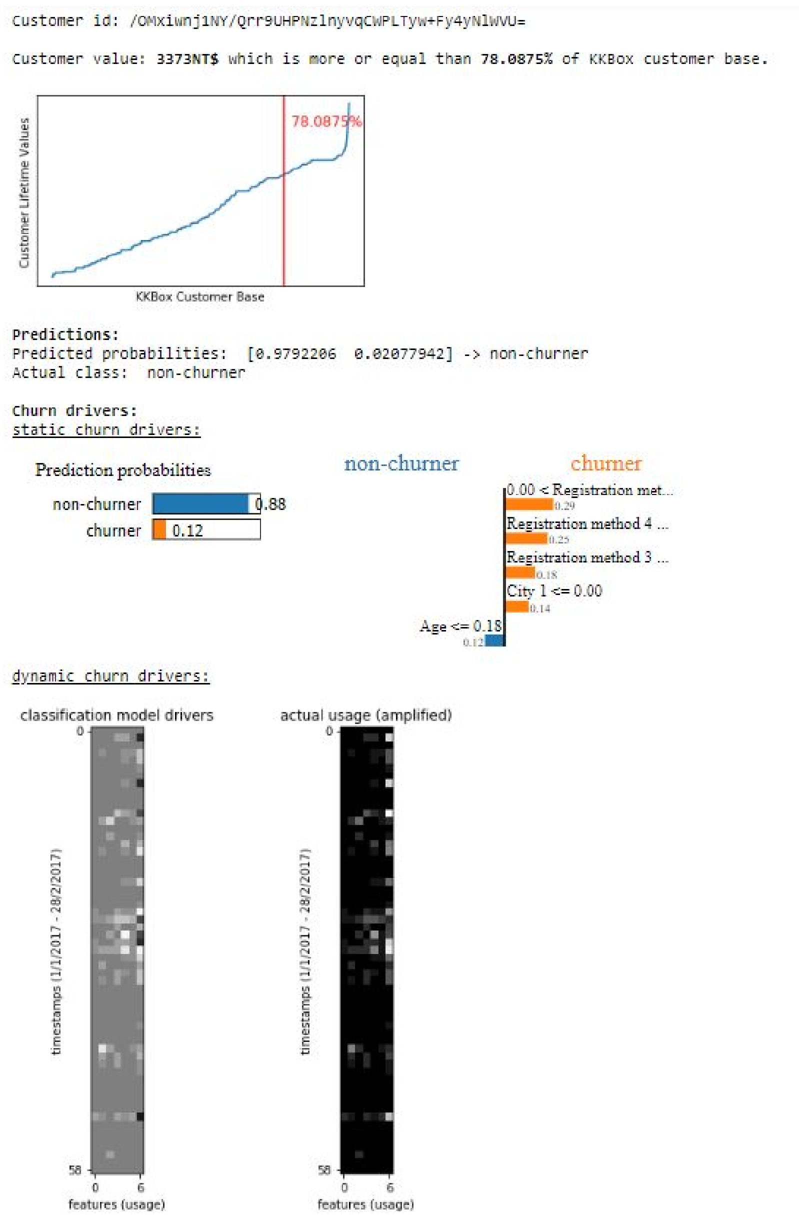


Figure 56: Managerial dashboard for an individual KKBox customer (non-churning customer)

Part 1: Customer Lifetime value
The CLV of the individual customer in this dashboard is 3373NT$ (the amount of money he or she paid throughout the whole tenure with KKBox). That means that the customer has paid more than 78,09% of the KKBox customer base.

Part 2: Churn probability
The prediction model favored non-churn class with 97,92% probability (the classification was correct).

Part 3: Churn predictors
In this particular case, the following features were in favor of the churning class:

- Registration method 9
- Registration method 4
- Registration method 3
- City 1

However, Age and the remaining features overweighted the decision in favor of non-churning.

As for the dynamic features, meaning customer behavior, the model once again does not provide any visual clues.

Discussion

The above-mentioned analysis shows how Deep Learning can provide a subscription-based company with highly accurate customer churn model predicting the probability of defection of each individual customer. Furthermore, by applying the interpretation model LIME model, it is possible to understand which predictors (features) contributed the most to the churn prediction of each customer. This information can provide valuable insight and information in several aspects of designing CRM retention strategy, as shown in the *Analysis* chapter.

The results show that the model is capable of predicting customer churn with high accuracy which not only brings the selection of key customers to the individual level, but also increases the probability of targeting the truly churning customers. As the literature shows, the increase of accuracy of a customer churn prediction has positive effect on the profits of retention campaigns, therefore using a highly accurate model of Deep Learning can be beneficial. However, the findings from the literature review also show that in order to increase the probability of a retention campaign, the selection of customers has to be based on several other factors as well. Therefore, the priority of companies should be also in the incorporation of these factors and not making the targeting decision only by the churn probabilities of the customers.

That being said, in regard to the subscription-based companies, the decision makers should be aware of the industry and the business model of the organization. The majority of the available research is conducted in the telecommunications industry (which is similar to subscription-based companies, available-data-wise), where differences between CLVs of respective customers can be significantly higher than in other subscription-based businesses. Therefore, in subscription-based companies customer's value might not be as significant of an aspect as the other proposed factors to the marketing interventions of the company. The inclusion of customer value also raises the question what are the retention goals which company aims to achieve. Nevertheless, it is important to note that the presented approach is based on maximizing profits of retention campaigns and includes calculating of customers' lifetime values. Additionally, the selection of customers is based on maximizing financial gains from the customer retention campaign, not ultimately minimizing the company's customer churn. On the other hand, the second approach of targeting customers only basis the selection of customers on their probability to response. The results of the study by Ascarza et al. (2018) show that this approach is more effective when a company aims to maximize the retention of customers. This approach may not be the most profitable in the short-term, however, it may be used in certain scenarios such as when a company aims to increase CLV of less profitable customers and turn them into more profitable ones.

Apart from the prediction of churn on individual level, the Deep Learning Customer Churn Prediction model can also yield the interpretations of the predictions' outcome, where the marketers are able to see which churn predictors are the most impactful in the calculated prediction. However, as described in the *Analysis* chapter, using a more complex Deep Learning model can lead to inability or limitations of using such an interpretation model. At that point, a company must decide on their priorities: whether they strive to use the most accurate model or if they are willing to accept the trade-off with clearer interpretations at the expense of predictions accuracy. These eventual predictors provide additional information about the drivers behind customer churn. Nevertheless, it is crucial to understand that these predictors of churn are not necessarily the reasons why customers defect. In order to utilize the information provided by the model, it is requisite for a company to understand what are the reasons behind churn of their customers in general and subsequently compare which one of those match with the predictors of churn for each individual customer provided by the model. Moreover, in order to successfully target the reasons of customer churn, companies have to distinguish between controllable and uncontrollable reasons of customer churn. Afterwards, the predictors of customers churn, which overlap with the controllable reasons of customer churn, can be utilized as marketing insights in retention efforts and actions. Lastly, it is noteworthy that understanding the reasons behind customer churn on the individual level is a difficult task. The current research of this area requires more exploration, which shall provide practical solutions to finding the drivers of churn for each customer. Nonetheless, if a company is unable to identify the reasons behind churn on the individual level, the information about predictors of churn can be combined with general drivers of churn, or the churn drivers of certain segments.

On top of affecting the targeting customers and reasons of customer churn, the data provided by customer churn prediction influences the development of CRM retention strategy itself. By being able to acquire new information about the customers on the individual level, the transition towards the individualized CRM is more feasible. In the case that a company already adapts the approach of individualized CRM strategy, the model provides new information to get a more complete picture about each individual customer. This enables companies to develop strategies toward each customer separately as well as to build relationships on the individual level. The individualized approach also allows companies to create tailor-made offers and messages, which are able to precisely satisfy customers' needs and strengthen the relationship between a company and a customer.

In order for company to use the Deep Learning method of customer churn prediction, several factors have to be evaluated. Firstly, a company has to look at the complexity of the available data it possesses. If a company aims to predict the customer churn by using complex data of different structure, the Deep Learning method may be the only option that is able to predict the customer churn. On the other side, if the available data is less complex, company may choose a simpler method of customer churn prediction. This may be a better decision due to the trade-off between the accuracy and the interpretation of the model. Furthermore, to use a simpler method may be less costly for a company, due to the challenging nature of creating a successful, highly accurate Deep Learning model. This is, however, closely connected to the context of the company and if there are any methods already used to predict customer churn. If a company uses less accurate Machine Learning model, but the interpretability of the model is reliable, the use of Deep Learning technique may not be beneficial. Especially, when the actions targeted on the reasons of churning are supported by the specific information provided by the interpretation of the model. On the other hand, if a company does not have any method to predicting customer churn, the mentioned advantages and disadvantages of the Deep Learning have to be evaluated in order to implement this technique. Nevertheless, the Deep Learning technique can improve the profits from retention efforts for companies who have already well-structured customer retention strategies. If the targeting of customers takes into consideration other important factors (expected CLV, costs, response probability and costs incentive), the increased accuracy of churn probability provided by Deep Learning increases the profit of the retention efforts.

As mentioned before, in order to create a complete fully informed CRM retention strategy, it is imperative to follow several interconnected steps. Should a company, however, utilize the Deep Learning standalone without taking into consideration the other steps required for CRM retention strategies, it might get detached from the bigger picture and overall goals.

This may, first and foremost, result in disharmony between different departments of the company. For example, if the strategic goal of the retention efforts is to maximize the ROI of the campaigns, the customer care department should be informed and trained in line with these goals in order to direct their retention efforts onto the right customers. Moreover, similarly to any other marketing efforts, the assessment division allowing company to reflect on the past efforts, examine the results and adjust if needed, is a crucial, yet often ignored part in the CRM retention strategy development. Companies ought to keep record of all the retention efforts, their success and financial return. Only under these conditions should further budgets be devoted to following retention endeavours.

Lastly, if the retention goals need to be in line with the internal and external context of the company. For example, in regard to the branding of a company, this may result in damaged brand image and in developed unfavorable customer behavior towards the brand. If a company strives to be seen as a serious and solid brand, unnecessary amount of customer discounts might create undesirable effects in the brand perception. This hypothesis, however, raises another question and that is whether companies outside of telecommunications industry and subscription-based model could enhance their CRM retention efforts with Deep Learning.

As examined thoroughly in the *Analysis* chapter, having the correct data is the single most important thing when attempting to develop a Deep Learning classification model. With the goal of identifying churning customers, companies ought to collect, store and process multitudes of data with regard to their customers. However, the research confirms that having static demographic information is not sufficient to predict churning customer with high accuracy. The most contemporary approaches employ behavioral data of their consumers either standalone or in combination with the static data. It is for this reason that customer churn detection methods heavily favor telecommunication or subscription-based companies in comparison to other companies. The firms in the two industries have at their disposal vast volumes of data of this nature in form of records of incoming or outgoing calls, SMS sent and received, data usage as well as number of songs listened or movies watched. It is therefore questionable whether companies from other industries such as banking or e-commerce can put together the data of this nature. Also, it is worth examining whether customer retention is of such an affectability as in the two industries discussed in this thesis. Or if customer acquisition and cross-sell are simply the sole most important aspects of their Customer Relationship Management.

Conclusion

Although the customer churn prediction has been of great interest of the telecommunication industry and companies with a subscription-based business model, there is a lack of academic research in regard to understanding how the concept of a customer churn prediction can fit into the processes of designing CRM retention strategies. Furthermore, due to the emergence of the Deep Learning technique only in the recent years, the research does not provide enough information about the marketing insights the Deep Learning Customer Churn Prediction can bring to the company. It is for this reason the researchers opted for the theme as a topic of the thesis. At the outset of this thesis, the research focuses on answering the following questions:

1. How to understand the relation between customer churn prediction and CRM retention strategy?
2. What are the most prominent churn drivers and which customers is it beneficial to intervene?
3. How can subscription-based companies utilize Deep Learning in customer churn prediction?
4. Can Deep Learning Customer Churn Prediction enhance traditional methods of CRM retention strategy?

In the following chapter the authors will conclude and summarize the findings related to the above-mentioned research objectives of this report. The first objective of the thesis was to understand customer churn prediction in relation to CRM retention strategy. To answer this question the researchers conducted a literature review of designing CRM retention strategy in order to understand the processes and aspects of this concept and recognize to which sections of the CRM retention strategy the customer churn prediction is related. By conceptualizing the existing literature review of CRM retention strategy, researchers are able to present a framework for developing and implementing CRM retention strategy. In the overall seven-step CRM retention strategy framework, customer churn prediction model is typically used in the process of selecting customers for the retention efforts but has an indirect influence on other parts as well.

One of them being the customers' reasons to churn whose understanding is the second objective of the thesis. The literature of this area most of the times agrees on the satisfaction, price, quality and complexity of offered service to be the most common reasons behind churning of customers in the telecommunications industry. However, the ability to distinguish between the controllable and uncontrollable reasons behind customer churn is an important factor, if a company aims to use these reasons for marketing interventions. In addition to the reasons of churn, the literature also describes the predictors of churn, which are the variables based on which a prediction model labels the customers as a churner or a non-churner. Furthermore, literature shows the connection of predictors of churn to the drivers of churn, where the predictors of churn are not necessarily the reasons behind customer churn but could match with them. In order for a company to utilize the insights from the predictors of churn for marketing use, the overlap between the predictors of churn and drivers of churn needs to be found.

The second part of this objective was to understand which customers are the most beneficial for a company to target in a retention strategy. The literature review of this area was conducted where multiple sources show that targeting based only on predictions of customer churn is not the most beneficial approach. If a company aims to maximize profits by customer retention efforts, it is necessary to base the selection of customers on the probability of customer churn combined with their expected CLV of customers, costs of incentives, and the probability of customers to response to the marketing actions. However, if company aims to maximize customer retention regardless of profits, the approach of targeting customers only by their probability to response may be the most effective approach.

In order unravel how can a subscription-based company utilize the Deep Learning Customer Churn Prediction model, the theoretical foundation for developing a Deep Learning model were studied and afterwards implemented in a real-market case scenario, where Deep Learning

Customer Churn Prediction model is constructed. It is clear both from the Theoretical background chapter, and from the Analysis chapter that adopting Deep Learning Customer Churn Prediction model in a subscription-based company is a challenging process comprising many interconnected steps. The most important prerequisite for a company trying to launch Deep Learning with the aim of predicting churning customers is having the proper data. It is for this reason that the companies with the most advanced churn predicting models come from telecommunications industry or subscription-based business model. These firms have at their hand vast amounts of customer data not only covering demographic information but also behavioral and transactional features. Proper data examination and preprocessing including feature engineering and understanding the nature of the data and its capabilities are the next crucial steps in the process. Getting the grasp of the nature of the available data is particularly important step as it determines the next step which is the actual building of the Deep Learning architecture. Once the model is built, the data is used for training and validation of the performance. Several metrics serve as an evaluation not only of the straightforward amount of correctly and incorrectly classified users, but also of each class' (churning and non-churning) results. Having this in place, the company can with high accuracy predict churning probability for each customer individually. On top of that, the company can utilize an interpretation model, more specifically LIME, and examine what features (meaning what input variables) contributed to the final probability with their respective weights. These churn predictors can be later used to derived further marketing insights. Nevertheless, the employment of the interpretability model come with certain architecture-specific issues. Should the Deep Learning Customer Churn Prediction model be too complex, it might put an obstacle in the attempts to interpret the results. At that point, the company needs to settle on the trade-off between the accuracy and interpretability when one comes at the expense of another.

Lastly, the findings from the constructed model are connected with the theoretical knowledge of Deep Learning modeling and CRM retention strategy. By doing so, the question of how can the Deep Learning Customer Churn Prediction model enhance the process of CRM retention strategy is presented and discussed. The findings indicate that Deep Learning Customer Churn Prediction model can be an extremely beneficial tool in the context of designing CRM retention strategies. As a matter of fact, it directly contributes to the part of the Customer Churn Insights where, as mentioned previously, it provides a set of churn predictors and their respective weights used for the predictions. These pieces of information in combination with controllable churn reasons can provide marketers with useful insights as to why an individual customer decided to leave the company. These can be later used to create a tailor-made offer which presumably increases the chances of the customer's reaction to intervention actions. In terms of indirect influence on the CRM retention strategies, the Deep Learning Customer Churn Prediction model affects the selection of key customers which is the preceding step of the abovementioned customer intervention. This step, though affecting the framework only indirectly, is extremely important as it can accurately predict which particular customers will likely leave the company. As a result, a list of likely-churning customers in combination with other factors such as Customer Lifetime Value or proneness to customer interventions, can significantly increase the ROI of retention strategies as they are focusing only on the truly key customers. At the same time, thanks to the ability to predict churning probability for each customer individually, the Deep Learning model brings the Degree of Individualized CRM to the individual level. Despite these advantages, it is imperative to maintain this model in the context of the entire development process as multitude of other factors come into picture when designing CRM retention strategies. These factors firstly include business context where internal and external factors influence the direction of the company and consequently their retention strategies. Furthermore, enabling process representing the execution of customer interventions as well as instruction of employees in the retention goals have great impact on the shape of the entire CRM retention strategy. And lastly, similarly to any marketing efforts, the results of the strategies ought to be recorded and evaluated.

Future research

As concluded in the Discussion, the conceptual framework described in chapter *Theoretical conceptualization of designing CRM retention strategies* enhanced by the Deep Learning Customer Churn Prediction model could yield profitable results. However, the presented approach illustrated in the conceptual framework is not fully applied in a real-market case scenario, due to the limited data used in the thesis. Therefore, further research can focus on the implementation of the entire CRM retention strategy strategy in order to fully evaluate the effectiveness of implementing Deep Learning Customer Churn Prediction model into this context. This would also lead to the possible comparison, whether customers who were classified as churners and consequently targeted through retention campaigns were actually going to churn. Furthermore, when the decision about which customers should be targeted is being made, the factor of considering a social network of a company has been mentioned by several studies. This means that churning of a certain customer may be also affected by churning of another customer, when these two customers are socially connected. Due to the inability to use this information in the already existing solutions to targeting customers, this factor was not included in this study. However, future research of selecting customers for retention may focus on implementing this factor into the process of more beneficial targeting of customers. Lastly, the future research can aim to provide more reliable and practical solutions to connecting predictors of customer churn with the reasons behind customer churn.

Limitations

First and foremost, due to the fact that the researchers were unable to close a collaboration with a company, the process of data-fetching which was essential for the Analysis part, outright restrained them from covering more of the individual fundamental parts of the Conceptual Framework. This way the calculation of Customer Lifetime Value could not meet the standards set by Lemmens et al. (2017) presented as an ideal case scenario for selecting key customers for retention efforts. Furthermore, no external and internal business factors are included. Lastly, nearly none of the churn predictors derived from the predictions could be simultaneously a churn reason. This creates an obstacle when fitting appropriate intervention actions on the predictions.

Next, although the researchers received an in-depth training in Python and Deep Learning, some of the advanced mathematical equations and explanations related to Neural Networks and its interpretations became fairly limiting when it comes to creating the most advanced model. On top of that, as the researchers ran into the limitations of the researched field regarding the interpretation of a multi-input Deep Learning architecture, the lack of necessary programming knowledge prevented them from finding a more sophisticated solution than the one presented in Analysis.

Lastly, another limiting factor came in the way of creating even more accurate prediction model. The computing power required for training of a Deep Learning architecture is immense and the more complex the architecture is, the more time- or power-consuming the process in. This way, the researchers could not explore behavioral patterns more in-depth. For example, to see the traits in all user behavioral features across a 7-day period on top of the current settings - one behavioral feature across 7 days and 7 behavioral features in one day.

Sources

Books and Journals

Abbasoğlu, M., Gedik, B. and Ferhatosmanoğlu, H. (2013). Aggregate profile clustering for telco analytics. Proceedings of the VLDB Endowment, 6(12), pp.1234-1237.

Ahmad, R. and Buttle, F. (2001). Customer retention: a potentially potent marketing management strategy. Journal of Strategic Marketing, 9(1), pp.29-45.

Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J. and Anwar, S. (2019). Customer churn prediction in telecommunication industry using data certainty. Journal of Business Research, 94, pp.290-301.

Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A. and Huang, K. (2017). Customer churn prediction in the telecommunication sector using a rough set approach. Neurocomputing, 237, pp.242-254.

Amin, A., Shah, B., Khattak, A., Lopes Moreira, F., Ali, G., Rocha, A. and Anwar, S. (2018). Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods. International Journal of Information Management.

Ang, L. and Buttle, F. (2006). Customer retention management processes. European Journal of Marketing, 40(1/2), pp.83-99.

Ansari, Asim and Carl F. Mela. (2003). "E-Customization," Journal of Marketing Research 40(2), 131-145.

Ansari, Asim, Carl Mela, and Scott A. Neslin. (2004). "Customer Channel Migration," Working Paper, Columbia University School of Business, NY.

Arora D, Malik P (2015) Analytics: key to go from generating big data to deriving business value. In: 2015 IEEE first international conference on big data computing service and applications, pp 446–452

Ascarza, E. (2018). Retention Futility: Targeting High-Risk Customers Might be Ineffective. Journal of Marketing Research, 55(1), pp.80-98.

Ascarza, E., Fader, P. and Hardie, B. (2017). Marketing Models for the Customer-Centric Firm.

Ascarza, E., Neslin, S., Netzer, O., Anderson, Z., Fader, P., Gupta, S., Hardie, B., Lemmens, A., Libai, B., Neal, D., Provost, F. and Schrift, R. (2017). In Pursuit of Enhanced Customer Retention Management: Review, Key Issues, and Future Directions. Customer Needs and Solutions, 5(1-2), pp.65-81.

Bengio, Y., 2009. Learning deep architectures for AI. Foundations and trends® in Machine Learning, 2(1), pp.1-127.

Bengio, Y., Courville, A. and Vincent, P., 2013. Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence, 35(8), pp.1798-1828.

Blattberg, R., Kim, B. and Neslin, S. (2008). Database Marketing. New York, NY: Springer-Verlag New York.

Braun, M. and Schweidel, D. (2011). Modeling Customer Lifetimes with Multiple Causes of Churn. Marketing Science, 30(5), pp.881-902.

Bryman, A. & Bell, E. (2011): Business Research Methods, 3rd edn. New York: Oxford University Press Inc., USA.

Bryman, A. (1988): Quantity and Quality in Social Research. London: Routledge.

Burrell, G. & Morgan, G. (1979): Sociological Paradigms and Organisational Analysis. Burlingtong: Ashgate Publishing Company, USA.

Buttle, F., 2004. Customer relationship management. Routledge.

Coviello, N., Brodie, R., Danaher, P. and Johnston, W. (2002). How Firms Relate to Their Markets: An Empirical Examination of Contemporary Marketing Practices. Journal of Marketing, 66(3), pp.33-46.

Dawkins, P. and Reichheld, F., 1990. Customer retention as a competitive weapon. Directors and Boards, 14(4), pp.42-47.

De Caigny, A., Coussement, K. and De Bock, K. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. European Journal of Operational Research, 269(2), pp.760-772.

Deng, L., Hinton, G. and Kingsbury, B., 2013, May. New types of deep neural network learning for speech recognition and related applications: An overview. In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 8599-8603). IEEE.

DeSouza, G. (1992). Designing a Customer Retention Plan. Journal of Business Strategy, 13(2), pp.24-28.

Dias, M., Locher, D., Li, M., El-Deredy, W. and Lisboa, P. (2008). The value of personalised recommender systems to e-business. Proceedings of the 2008 ACM conference on Recommender systems - RecSys '08.

Esghi, A., Haughton, D. and Topi, H. (2007). Determinants of customer loyalty in the wireless telecommunications industry. Telecommunications Policy, 31(2), pp.93-106.

F. Castanedo, G. Valverde, J. Zaratiegui and A. Vazquez, Using Deep Learning to Predict Customer Churn in a Mobile Telecommunication Net- works, Wise Athena LLC, 2014. [Online]. Available: http://wiseathena. com/pdf/wa dl.pdf [Accessed 01 05 2015].

Fan, S., Lau, R. and Zhao, J. (2015). Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix. Big Data Research, 2(1), pp.28-32.

Fernández-García, A., Iribarne, L., Corral, A., Criado, J. and Wang, J. (2019). A recommender system for component-based applications using Machine Learning techniques. Knowledge-Based Systems, 164, pp.68-84.

Foltean, F., Trif, S. and Tuleu, D. (2018). Customer relationship management capabilities and social media technology use: Consequences on firm performance. Journal of Business Research.

Frow, P. and Payne, A. (2009). Customer Relationship Management: A Strategic Perspective. Journal of business market management, 3(1), pp.7-27.

Gerpott, T., Rams, W. and Schindler, A. (2001). Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. Telecommunications Policy, 25(4), pp.249-269.

Gordini, N. and Veglio, V. (2019). Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry.

Grönroos, C. (1991). The Marketing Strategy Continuum: Towards a Marketing Concept for the 1990s. Management Decision, 29(1).

Guelman, L., Guillén, M. and Pérez-Marín, A. (2012). 'Random Forests for Uplift Modeling: An Insurance Customer Retention Case', in L., Engemann, K., Gil-Lafuente, A. and Merigó, J. (2012). Modeling and Simulation in Engineering, Economics and Management. Berlin, Heidelberg: Springer Berlin Heidelberg.

Guelman, L., Guillén, M. and Pérez-Marín, A. (2015). Uplift Random Forests. Cybernetics and Systems, 46(3-4), pp.230-248.

Gui, C. (2017). Analysis of imbalanced data set problem: The case of churn prediction for telecommunication. Artificial Intelligence Research, 6(2), p.93.

Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., Ravishanker, N. and Sriram, S. (2006). Modeling Customer Lifetime Value. Journal of Service Research, 9(2), pp.139-155.

Gustafsson, A., Johnson, M. and Roos, I. (2005). The Effects of Customer Satisfaction, Relationship Commitment Dimensions, and Triggers on Customer Retention. Journal of Marketing, 69(4), pp.210-218.

Gustafsson, A., Johnson, M. D., & Roos, I. (2005). The effects of customer satisfaction, relationship commitment dimension, and triggers on customer retention. Journal of Marketing, 69(4), 210–218.

Hadden, J., Tiwari, A., Roy, R. and Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. Computers & Operations Research, 34(10), pp.2902-2917.

Hedin, H., Hirvensalo, I. and Vaarnas, M. (2014). The Handbook of Market Intelligence: Understand, Compete and Grow in Global. John Wiley & Sons.

Jyh‐Fu Jeng, D. and Bailey, T. (2012). Assessing customer retention strategies in mobile telecommunications. Management Decision, 50(9), pp.1570-1595.

Kamakura, W. and Wedel, M. (2003). List augmentation with model based multiple imputation: a case study using a mixed-outcome factor model. Statistica Neerlandica, 57(1), pp.46-57.

Kamakura, W., Mela, C., Ansari, A., Bodapati, A., Fader, P., Iyengar, R., Naik, P., Neslin, S., Sun, B., Verhoef, P., Wedel, M. and Wilcox, R. (2005). Choice Models and Customer Relationship Management. Marketing Letters, 16(3-4), pp.279-291.

Kamakura, W., Ramaswami, S. and Srivastava, R. (1991). Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. International Journal of Research in Marketing, 8(4), pp.329-349.

Keaveney, S. (1995). Customer Switching Behavior in Service Industries: An Exploratory Study. Journal of Marketing, 59(2), p.71.

Keramati, A., Mehrabi, H., & Mousakhani, M. (2008). The role of human, organization and technology in building competitive advantage with CRM: The case of iranian

Ketkar, N. (2017). Deep Learning with Python. Berkeley, CA: Apress.

Ko, E., Kim, S. H., Kim, M., & Woo, J. Y. (2008). Organizational characteristics and the CRM adoption process. Journal of Business Research, 61(1), 65–74.

Kuada, J. (2012). Research methodology. Frederiksberg C [Denmark]: Samfundslitteratur.

Kuhn, T. (1970). The structure of scientific revolutions. Second edition, enlarged. [Chicago, London]: [] University of Chicago Press.

Kumar, V. and Petersen, J. (2005). Using a Customer-Level Marketing Strategy to Enhance Firm Performance: A Review of Theoretical and Empirical Evidence. Journal of the Academy of Marketing Science, 33(4), pp.504-519.

LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep Learning. Nature, 521(7553), pp.436-444.

Lee, B., Collier, P. and Cullen, J. (2007). Reflections on the use of case studies in the accounting, management and organizational disciplines. Qualitative Research in Organizations and Management: An International Journal, 2(3), pp.169-178.

Lemmens, A. and Gupta, S. (2013). Managing Churn to Maximize Profits. Working Paper, no. 14-020, Harvard Business School, Boston.

Lemmens, A. and Gupta, S. (2017). Managing Churn to Maximize Profits. SSRN Electronic Journal.

Li, Y., Jiang, W., Yang, L. and Wu, T. (2018). On neural networks and learning systems for business computing. Neurocomputing, 275, pp.1150-1159.

Lindgreen, A., Davis, R., Brodie, R.J. and Buchanan-Oliver, M., 2000. Pluralism in contemporary marketing practices. International Journal of Bank Marketing, 18(6), pp.294-308.

Mills, S., Lucas, S., Irakliotis, L., Rappa, M., Carlson, T. and Perlowitz, B., 2012. Demystifying big data: a practical guide to transforming the business of government. TechAmerica Foundation, Washington.

Montavon et al. (2018). Methods for interpreting and understanding deep neural networks. Digital Signal Processing, 73, pp.1-15.

Montavon, G., Lapuschkin, S., Binder, A., Samek, W. and Müller, K. (2017). Explaining nonlinear classification decisions with deep Taylor decomposition. Pattern Recognition, 65, pp.211-222.

Morgan, N. A., Slotegraaf, R. J., & Vorhies, D. W. (2009). Linking marketing capabilities with profit growth. International Journal of Research in Marketing, 26(4), 284–293.

Neslin, S., Gupta, S., Kamakura, W., Lu, J. and Mason, C. (2006). Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models. Journal of Marketing Research, 43(2), pp.204-211.

Payne, A. and Frow, P. (1999). Developing a Segmented Service Strategy: Improving Measurement in Relationship Marketing. Journal of Marketing Management, 15(8), pp.797-818.

Payne, A. and Frow, P. (2005). A Strategic Framework for Customer Relationship Management. Journal of Marketing, 69(4), pp.167-176.

Payne, A. and Frow, P. (2006). Customer Relationship Management: from Strategy to Implementation. Journal of Marketing Management, 22(1-2), pp.135-168.

Putter, H., Fiocco, M. and Geskus, R. (2007). Tutorial in biostatistics: competing risks and multi-state models. Statistics in Medicine, 26(11), pp.2389-2430.

Ramsundar, R. (2018). TensorFlow for Deep Learning. [S.l.]: O'Reilly Media, Inc.

Reichheld, F. (1996). The loyalty effect: The hidden force behind growth, profits, and lasting value. Long Range Planning, 29(6), p.909.

Reichheld, F. and Sasser, Jr., W. (1990). Zero Defections: Quality Comes to Services. [online] Harvard Business Review. Available at: https://hbr.org/1990/09/zero-defections-quality-comes-to-services [Accessed 5 Jun. 2019].

Ribeiro et al. (2016). "Why Should I Trust You?".

Rosenberg, L. and Czepiel, J. (1984). A MARKETING APPROACH FOR CUSTOMER RETENTION. Journal of Consumer Marketing, 1(2), pp.45-51.

Ryals, L. (2005). Making Customer Relationship Management Work: The Measurement and Profitable Management of Customer Relationships. Journal of Marketing, 69(4), pp.252-261.

Seo, D., Ranganathan, C. and Babad, Y. (2008). Two-level model of customer retention in the US mobile telecommunications service market. Telecommunications Policy, 32(3-4), pp.182-196.

Sharp, B. (2010). How brands grow. Oxford University Press, p.29.

Shrikumar, A., Greenside, P. and Kundaje, A., 2017, August. Learning important features through propagating activation differences. In Proceedings of the 34th International Conference on Machine Learning-Volume 70 (pp. 3145-3153). JMLR. org.

Siddaway, A. (2014). What is a systematic literature review and how do I do one. University of Stirling, (I), p.1.

Silverman, D. (1985). Qualitative Methodology and Sociology: Describing the Social World. Aldershot: Gower.

Snape, D. & Spencer, L. (2003). 'The Foundations of Qualitative Research', in Ritchie, J., & Lewis, J. (2003). Qualitative research practice: a guide for social science students and researchers. London, Sage Publications.

Soltani, Z., Zareie, B., Milani, F. and Navimipour, N. (2018). The impact of the customer relationship management on the organization performance. The Journal of High Technology Management Research, 29(2), pp.237-246.

Spanoudes, P. and Nguyen, T., 2017. Deep learning in customer churn prediction: unsupervised feature learning on abstract company independent feature vectors. arXiv preprint arXiv:1703.03869.

Tranfield, D., Denyer, D. and Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. British Journal of Management, 14(3), pp.207-222.

Verbeke, W., Dejaeger, K., Martens, D., Hur, J. and Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. European Journal of Operational Research, 218(1), pp.211-229.

Wang, J. and Song, G. (2018). A Deep Spatial-Temporal Ensemble Model for Air Quality Prediction. Neurocomputing, 314, pp.198-206.

Wang, Y., & Feng, H. (2012). CRM capabilities. Management Decision, 50, 115–129.

Wangperawong, A., Brun, C., Laudy, O. and Pavasuthipaisit, R., 2016. Churn analysis using deep convolutional neural networks and autoencoders. arXiv preprint arXiv:1604.05377.

Weinstein, A. (2002). Customer retention: A usage segmentation and customer value approach. Journal of Targeting, Measurement and Analysis for Marketing, 10(3), pp.259-268.

Xevelonakis, E. (2005). Developing retention strategies based on customer profitability in telecommunications: An empirical study. Journal of Database Marketing & Customer Strategy Management, 12(3), pp.226-242.

Yim, F. H., Anderson, R. E., and Swaminathan, S., (2005). Customer Relationship Management: Its Dimensions and Effect on Customer Outcomes. Journal of Personal Selling and Sales Management, 24(4), 265-280.

Yin, R. (1994). Case study research. Newbury Park, Calif.: Sage Publications.

Yin, R. (2003). Case study research. Thousand Oaks, Calif.: SAGE Publications.

Zhang, M., Guo, L., Hu, M. and Liu, W., 2017. Influence of customer engagement with company social networks on stickiness: Mediating effect of customer value creation. International Journal of Information Management, 37(3), pp.229-240.

Zineldin, M. (2006). The royalty of loyalty: CRM, quality and retention. Journal of Consumer Marketing, 23(7), pp.430-437.

Websites

Bronshtein, A. (2017). Train/Test Split and Cross Validation in Python. [online] Towards Data Science. Available at: https://towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61beca4b6 [Accessed 13 May 2019].

Brownlee, J. (2016). Display Deep Learning Model Training History in Keras. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/display-deep-learning-model-training-history-in-keras/ [Accessed 22 May 2019].

Brownlee, J. (2017). Data Preparation for Variable Length Input Sequences. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/data-preparation-variable-length-input-sequences-sequence-prediction/ [Accessed 13 May 2019].

Brownlee, J. (2017). How to Use the Keras Functional API for Deep Learning. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/keras-functional-api-deep-learning/ [Accessed 22 May 2019].

Brownlee, J. (2018). When to Use MLP, CNN, and RNN Neural Networks. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/when-to-use-mlp-cnn-and-rnn-neural-networks/ [Accessed 2 Jun. 2019].

Brownlee, J. (2019). A Gentle Introduction to Learning Curves for Diagnosing Machine Learning Model Performance. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance/ [Accessed 22 May 2019].

Cornelisse, D. (2018). An intuitive guide to Convolutional Neural Networks. [online] freeCodeCamp.org. Available at: https://medium.freecodecamp.org/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050 [Accessed 22 May 2019].

DeMuro, J. (2018). What is a neural network?. [online] TechRadar. Available at: https://www.techradar.com/news/what-is-a-neural-network [Accessed 5 Jun. 2019].

Donges, N. (2018). Recurrent Neural Networks and LSTM. [online] Towards Data Science. Available at: https://towardsdatascience.com/recurrent-neural-networks-and-lstm-4b601dd822a5 [Accessed 2 Jun. 2019].

Doukkali, F. (2017). Batch normalization in Neural Networks. [online] Towards Data Science. Available at: https://towardsdatascience.com/batch-normalization-in-neural-networks-1ac91516821c [Accessed 22 May 2019].

GitHub. (2019). LIME and Multi-input CNN model · Issue #317 · marcotcr/lime. [online] Available at: https://github.com/marcotcr/lime/issues/317 [Accessed 2 Jun. 2019].

Goodrum, W. (2016). Balance: Accuracy vs. Interpretability in Regulated Environments. [online] Elderresearch.com. Available at:

https://www.elderresearch.com/blog/predictive-model-accuracy-versus-interpretability [Accessed 22 May 2019].

Google Developers. (2019). Classification: Accuracy | Machine Learning Crash Course | Google Developers. [online] Available at: https://developers.google.com/machine-learning/crash-course/classification/accuracy [Accessed 5 Jun. 2019].

Joseph, J. (2016). How to treat missing values in your data : Part II. [online] CleverTap. Available at: https://clevertap.com/blog/how-to-treat-missing-values-in-your-data-part-ii/ [Accessed 13 May 2019].

Kaggle.com. (2017). WSDM - KKBox's Churn Prediction Challenge. [online] Available at: https://www.kaggle.com/c/kkbox-churn-prediction-challenge/data [Accessed 13 May 2019].

Kathuria, A. (2019). Methods and Uses of Feature Scaling. [online] Medium. Available at: https://medium.com/datadriveninvestor/methods-and-uses-of-feature-scaling-94a44b43181a [Accessed 13 May 2019].

Keras.io. (n.d.). Core Layers - Keras Documentation. [online] Available at: https://keras.io/layers/core/#dropout [Accessed 22 May 2019].

Koehrsen, W. (2018). How to Create Value with Machine Learning. [online] Towards Data Science. Available at: https://towardsdatascience.com/how-to-create-value-with-machine-learning-eb09585b332e [Accessed 13 May 2019].

Lundberg, S.M. and Lee, S.I., 2017. A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (pp. 4765-4774).

Maladkar, K. (2019). 6 Types of Artificial Neural Networks Currently Being Used in ML. [online] Analytics India Magazine. Available at: https://www.analyticsindiamag.com/6-types-of-artificial-neural-networks-currently-being-used-in-todays-technology/ [Accessed 2 Jun. 2019].

Malik, U. (2018). Time Series Analysis with LSTM using Python's Keras Library. [online] Stack Abuse. Available at: https://stackabuse.com/time-series-analysis-with-lstm-using-pythons-keras-library/ [Accessed 2 Jun. 2019].

Mehta, A. (2019). A Complete Guide to Types of Neural Networks. [online] Digital Vidya. Available at: https://www.digitalvidya.com/blog/types-of-neural-networks/ [Accessed 2 Jun. 2019].

Mishra, A. (2018). Metrics to Evaluate your Machine Learning Algorithm. [online] Towards Data Science. Available at: https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234 [Accessed 22 May 2019].

Narkhede, S. (2018). Understanding AUC - ROC Curve. [online] Towards Data Science. Available at: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5 [Accessed 22 May 2019].

Parmar, R. (2018). Common Loss functions in Machine Learning. [online] Towards Data Science. Available at: https://towardsdatascience.com/common-loss-functions-in-machine-learning-46af0ffc4d23 [Accessed 2 Jun. 2019].

Perez Denadai, E. (2018). Interpretability of Deep Learning Models. [online] Towards Data Science. Available at: https://towardsdatascience.com/interpretability-of-deep-learning-models-9f52e54d72ab [Accessed 22 May 2019].

Ping Shung, K. (2018). Accuracy, Precision, Recall or F1?. [online] Towards Data Science. Available at: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9 [Accessed 22 May 2019].

Raj, D. (2016). Learning Python Regression Analysis — part 7: Handling outliers in data. [online] Medium. Available at: https://medium.com/@dhwajraj/learning-python-regression-analysis-part-7-handling-outliers-in-data-d36ee9e2130b [Accessed 5 Jun. 2019].

Ray, S. (2016). A Complete Tutorial which teaches Data Exploration in detail. [online] Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2016/01/guide-data-exploration/ [Accessed 13 May 2019].

Robbins, N. (2012). When Should I Use Logarithmic Scales in My Charts and Graphs?. [online] Forbes.com. Available at: https://www.forbes.com/sites/naomirobbins/2012/01/19/when-should-i-use-logarithmic-scales-in-my-charts-and-graphs/#58dddb595e67 [Accessed 13 May 2019].

Ryckel, F. (2017). Logistic Regression Part 1. [online] Fderyckel.github.io. Available at: https://fderyckel.github.io/2017-03-15-Logistic_Regression_Part_I/ [Accessed 22 May 2019].

Sharma V, A. (2017). Understanding Activation Functions in Neural Networks. [online] Medium. Available at:

https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0 [Accessed 22 May 2019].

Sharma, M. (2018). What Steps should one take while doing Data Preprocessing?. [online] HackerNoon. Available at: https://hackernoon.com/what-steps-should-one-take-while-doing-data-preprocessing-502c993e1caa [Accessed 13 May 2019].

Sharma, S. (2017). Epoch vs Batch Size vs Iterations. [online] Towards Data Science. Available at: https://towardsdatascience.com/epoch-vs-iterations-vs-batch-size-4dfb9c7ce9c9 [Accessed 22 May 2019].

Shiebler, D. (2017). Understanding Neural Networks with Layerwise Relevance Propagation and Deep Taylor Series. [online] Danshiebler.com. Available at: http://danshiebler.com/2017-04-16-deep-taylor-lrp/ [Accessed 22 May 2019].

Tch, A. (2017). The mostly complete chart of Neural Networks, explained. [online] Towards Data Science. Available at: https://towardsdatascience.com/the-mostly-complete-chart-of-neural-networks-explained-3fb6f2367464 [Accessed 2 Jun. 2019].

Usmani, Z. (2017). What is Kaggle, Why I Participate, What is the Impact?. [online] Kaggle.com. Available at: https://www.kaggle.com/getting-started/44916 [Accessed 3 Jun. 2019].

Zhu, F., Zhong, C., Fang, S., Bouchard, R., Fontama, V., Song, X., Singh, P., Gao, J. and Deng, L. (2017). Predicting Azure Churn with Deep Learning and Explaining Predictions with LIME. In: Strata Hadoop. [online] Available at:

https://conferences.oreilly.com/strata/strata-ca-2017/public/schedule/detail/56135 [Accessed 20 Feb. 2019].

List of Figures

Figures

Tables

Appendix

APPENDIX 1.

Author	**Year**	**Name**	**Relation to the section of thesis**
Ahmad et al.	2001	Customer retention: a potentially potent marketing management strategy	CRM retention strategy
Ascarza	2018	Retention Futility: Targeting High-Risk Customers Might Be Ineffective	Selection of customers to target
Ascarza et al.	2017	In pursuit of enhanced customer retention management	CRM retention strategy, Selection of customers to target, Drivers of churn
Ascarza et al.	2017	Marketing Models for the Customer-Centric Firm	CRM retention strategy, Selection of customers to target, Drivers of churn
Bengio	2009	Learning deep architectures for AI. Foundations and trends® in Machine Learning	Deep Learning Neural Networks
Bengio et al.	2013	Representation learning: A review and new perspectives	Deep Learning Neural Networks
Blattberg et al.	2008	Database Marketing	Selection of customers to target
Braun et al.	2011	Modeling Customer Lifetimes with Multiple Causes of Churn	Drivers of churn
Bronshtein	2017	Train/Test Split and Cross Validation in Python	Train and Test split

Brownlee	2016	Display Deep Learning Model Training History in Keras	Training process
Brownlee	2018	When to Use MLP, CNN, and RNN Neural Networks	Types of Deep Learning Neural Networks
Cornelisse	2018	An intuitive guide to Convolutional Neural Networks	Churn analysis using multiple data sources
De Caigny et al.	2018	A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees	Deep Learning Customer Churn model
DeMuro	2018	What is a neural network?	Deep Learning Neural Networks
DeSouza	1992	Designing a Customer Retention Plan	Drivers of churn
Deng et al.	2013	New types of deep neural network learning for speech recognition and related applications: An overview	Deep Learning Neural Networks
Donges	2018	Recurrent Neural Networks and LSTM	Types of Deep Learning Neural Networks
Esghi	2007	Determinants of customer loyalty in the wireless telecommunications industry	Drivers of churn
Frow et al.	2009	Customer Relationship Management: A Strategic Perspective	CRM retention strategy
Gerpott et al.	2001	Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market	CRM retention strategy
Goodrum	2016	Balance: Accuracy vs. Interpretability in Regulated Environments	Interpretability of Deep Neural Networks
Guelman et al.	2012	Random Forests for Uplift Modeling: An Insurance Customer Retention Case	Selection of customers to target
Guelman et al.	2015	Uplift Random Forests	Selection of customers to target

Gui	2017	Analysis of imbalanced data set problem: The case of churn prediction for telecommunication	Imbalanced class distribution
Gustafsson	2005	The Effects of Customer Satisfaction, Relationship Commitment Dimensions, and Triggers on Customer Retention	Drivers of churn
Jeng et al.	2012	Assessing customer retention strategies in mobile telecommunications: Hybrid MCDM approach	Drivers of churn
Keaveney	1995	Customer Switching Behavior in Service Industries: An Exploratory Study	Drivers of churn
Kumar et al.	2005	Using a Customer-Level Marketing Strategy to Enhance Firm Performance: A Review of Theoretical and Empirical Evidence	CRM retention strategy
Lemmens et al.	2013	Managing Churn to Maximize Profits	Selection of customers to target
Lemmens et al.	2017	Managing Churn to Maximize Profits	Selection of customers to target
Lundberg et al.	2017	A unified approach to interpreting model predictions	Interpretability of Deep Neural Networks
Maladkar	2019	6 Types of Artificial Neural Networks Currently Being Used in ML	Types of Deep Learning Neural Networks
Malik	2018	Time Series Analysis with LSTM using Python's Keras Library	Types of Deep Learning Neural Networks
Mehta	2019	A Complete Guide to Types of Neural Networks	Types of Deep Learning Neural Networks
Montavon et al.	2018	Methods for interpreting and understanding deep neural networks	Interpretability of Deep Neural Networks

Montavon et al.	2017	Explaining nonlinear classification decisions with deep Taylor decomposition	Interpretability of Deep Neural Networks
Neslin et al.	2006	Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models	Selection of customers to target
Payne et al.	1999	Developing a Segmented Service Strategy: Improving Measurement in Relationship Marketing	CRM retention strategy
Payne et al.	2005	A Strategic Framework for Customer Relationship Management	CRM retention strategy
Payne et al.	2006	Customer Relationship Management: from Strategy to Implementation	CRM retention strategy
Perez Denadai	2018	Interpretability of Deep Learning Models	Interpretability of Deep Neural Networks
Putter et al.	2007	Tutorial in biostatistics: competing risks and multi-state models. Statistics in Medicine	Drivers of churn
Ramsundar	2018	TensorFlow for Deep Learning	Types of Deep Learning Neural Networks
Ribeiro et al.	2016	"Why Should I Trust You?"	Interpretability of Deep Neural Networks
Ryals	2005	Making Customer Relationship Management Work: The Measurement and Profitable	CRM retention strategy
Seo et al.	2008	Two-level model of customer retention in the US mobile telecommunications service market	Drivers of churn
Shiebler	2017	Understanding Neural Networks with Layerwise Relevance Propagation and Deep Taylor Series	Interpretability of Deep Neural Networks

Shrikumar et al.	2017	Learning important features through propagating activation differences	Interpretability of Deep Neural Networks
Spanoudes et al.	2017	Deep learning in customer churn prediction: unsupervised feature learning on abstract company independent feature vectors	Deep Learning Neural Networks
Tch	2017	The mostly complete chart of Neural Networks, explained	Types of Deep Learning Neural Networks
Verbeke	2012	New insights into churn prediction in the telecommunication sector: A profit driven data mining approach	Selection of customers to target, Drivers of churn
Verbeke et al.	2012	New insights into churn prediction in the telecommunication sector: A profit driven data mining approach	Deep Learning Customer Churn model
Wangperawong et al.	2016	Churn analysis using deep convolutional neural networks and autoencoders	Churn analysis using image recognition
Weinstein	2001	Customer-Specific Strategies Customer retention: A usage segmentation and customer value approach	CRM retention strategy
Xevelonakis	2005	Developing retention strategies based on customer profitability in telecommunications: An empirical study	CRM retention strategy
Yim et al.	2004	Customer Relationship Management: Its Dimensions and Effect on Customer Outcomes	CRM retention strategy
Zhu et al.	2017	Predicting Azure Churn with Deep Learning and Explaining Predictions with LIME	Churn analysis using multiple data sources
Zineldin	2006	The royalty of loyalty: CRM, quality and retention	CRM retention strategy